Drug Innovations and Welfare Measures Computed from Market Demand: The Case of Anti-Cholesterol Drugs*

Abe Dunn

October 19, 2010

Abstract

The pharmaceutical industry is characterized as having substantial investment in R&D and a large number of new product introductions, which poses special problems for price measurement caused by the quality of drug products changing over time. This paper applies recent demand estimation techniques to construct a constant-quality price index for anti-cholesterol drugs. Demand is estimated using a nationally representative sample of individuals over the period 1996 to 2007 that includes detailed information on individual health conditions, demographics, insurance, and prescription drug choices. Although the average price for anti-cholesterol drugs does not change over the sample period, I find that the constant-quality price index drops by 22 percent, a pace more in line with our expectations in such a dynamic segment of the industry. This result is robust to a number of alternative assumptions, highlighting the importance of controlling for quality in markets with significant innovation. The demand estimates also reveal that the benefits from new innovations depend on the health conditions of individuals which may impact quality-adjusted prices for different populations.

1 Introduction

The growth in medical technology is a driving force behind the rising costs of medical care. Based on studies by Newhouse (1992), Cutler (1995), and Smith et al (2000), new technologies account for approximately 50 percent of cost growth in medical care in recent decades.[1] Although new technologies often lead to higher expenditures on medical care, they also affect the quality of treatment, typically improving patient welfare and lowering the

*I would like to thank Ana Aizcorbe, Ralph Bradley, Gautum Gowrisankaran, John Greenlees, Chuck Romeo, Brett Wendling and seminar participants at the Bureau of Labor Statistics for comments. I would also like to thank Karen Rasmussen M.D. for sharing her knowledge about cholesterol treatment. This paper also benefited from comments by Randal Watson, Stephen Donald, and Ken Hendricks on an earlier version of this paper. Sarah Pack provided research assistance. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Bureau of Economic Analysis.

[1] The Congressional Budget Office (CBO (2008)) reviewed these studies and concluded the approximate effect of technology to be about 50% of the cost growth. A more recent study by Smith et al (2009) estimates that medical technology explains 27-48% of health spending growth since 1960.

quality-adjusted cost of treatment. The rapid shift in product quality over time poses special challenges for price measurement.

Price index estimates that hold quality fixed are critical for measuring real output in the healthcare sector and may also inform public policies related to innovation. If the price index falls as innovative products enter the market, this suggests that innovations have led to improved treatments, relative to the cost, and we should continue supporting policies that promote innovation. Conversely, if the price index increases when new products enter the market, then one might conclude that innovations, in some sense, were not worth the cost.

The pharmaceutical industry has been the most R&D intensive industry in the manufacturing sector for the past 20 years and is an areas of medical care where new technologies are prevalent.[2] Among pharmaceutical treatments, anti-cholesterol drugs are one of the most important areas of innovation based on their impact on health and innovations over the past three decades. Extensive medical evidence has shown that high cholesterol is a contributing factor in 56 percent of clinical heart disease cases, which is the leading cause of death in the United States. The introduction of the statin class of cholesterol-lowering drugs starting in 1987 has proven to be a critical development for preventing heart disease. Many individuals with high cholesterol can expect to gain many months or years of additional life by using statin treatments.[3] Innovations in this area have led to rapid growth, with the use of anti-cholesterol medications growing more than 400 percent over the period of study from 1996 to 2007.

This paper uses a demand model for anti-cholesterol drugs to construct a price index that accounts for quality changes resulting from new product introductions. The approach applied in this paper has been used to assess the value of new goods in a variety of industries.[4] However, relatively few papers have applied these techniques to examine the impact of innovations in the medical care sector. One of the earlier papers examining innovation in the health sector was Trajtenberg (1990) who examined innovation in the CT scanner market. More recent work has focused on the pharmaceutical industry with Cleanthous (2004) studying the innovation for depression drugs and Lucarelli and Nicholson (2009) looking at new colorectal cancer drugs.

The demand for anti-cholesterol drugs is modeled using a discrete-choice framework similar to Berry (1994) and BLP (1995). In contrast to the previous work examining innovation in the health sector that uses aggregate data, the model presented here uses detailed, nationally representative individual-level data that includes information on health conditions, demographics, health insurance, drug insurance, and individual-specific drug choices. The

[2] CBO (2006) reports pharmaceuticals as one of the most R&D intensive industries based on Pharmaceutical Research Manufacturing Association estimates. More conservative estimates from the National Science Foundation show that pharamaceuticals are the most R&D intensive industry for most of the past 20 years, but they find that the Communications Equipment sector exceeds the pharmaceutical sector in research intensity post-1999.

[3] The U.K. study by Ward et al (2007) provides a nice review of these studies and conducts a meta-analysis of the effectiveness of these drugs. The Heart Protection Collaboration Group (2010) also reviews the literature and conducts a meta-analysis using U.S. data and analyzes the effectiveness of Statins for people with different levels of cardiovascular risk. Even when one considers partial noncompliance with treatment and discounts future years of life gained, consumers may expect to gain several months in additional life for a relatively small cost (See Ward et al (2007)).

[4] For example, automobiles (Berry, Levinson, and Pakes (1993) , Petrin (2002)), computers (Greenstein (1994)), and breakfast cereals (Nevo (2003)). For a more complete review of the literature see Bresnahan and Gordon (1997).

model permits flexible substitution patterns that are affected by the observed health conditions and demographics of individuals in the market. This model is particularly well-suited for estimating the welfare for new medications since the effectiveness of drugs and their side effects may vary depending on the severity of the condition, the specifics of the disease, and the demographics of the individual. Using individual level information on drug insurance coverage I am also able to control for potential moral hazard effects that may distort the market valuation of anti-cholesterol drugs. Although choices are modeled using detailed individual level information, a key advantage of the model is that it is relatively simple to implement and may be applied to other drug classes with readily available healthcare databases.

The results indicate that the quality-adjusted price of anti-cholesterol drugs has fallen considerably since 1996, reflecting the importance of innovation in this market. Relative to the CPI, the quality-adjusted price *fell* by 5 percent from 1996 to 2005, while the average price *grew* by 37 percent. Both average prices and real prices fell sharply after 2005 following the entry of generics. While the quality-adjusted index shows a decline in the real price of anti-cholesterol drug treatment, the quality of a new treatment may depend on the health condition of the patient. More precisely, the individual utility from a new drug is derived through the expected treatment of a condition, rather than the treatment itself, so that the impact of an innovation will be unique to each individual and her condition. This paper presents a methodology for constructing a condition-specific price index for pharmaceutical products. I examine the hypothetical impact of introducing statins in the market on quality-adjusted prices for those with heart disease and those without to show how innovations impact different populations. Conservative estimates suggest that for those without heart disease the introducing statins in the market in 2007 would be equivalent to a 26.5 percent reduction in quality-adjusted price, while the price reduction for those with heart disease would be 30.5 percent.

This paper also addresses some critical assumptions that must be made when constructing price indices from demand estimates. Specifically, when looking at breakfast cereals, Nevo (2003) finds that the changes in quality-adjusted price indices may hinge on assumptions regarding demand trends and unobservable demand characteristics. In particular, he finds that price indices derived from estimated demand systems may change significantly depending on whether a researcher treats unobservable demand and trends as changes in product attributes or shifts in individual tastes. These assumptions are important because if unobserved demand and trends capture changes in individual taste, they should not be viewed as shifts in quality and should be held fixed over time. Similar to Nevo, I find differences among quality-adjusted price indices depending on how one views unobserved demand and trends. In contrast to Nevo, I find that in all cases the quality-adjusted prices are much lower than the average price; while Nevo finds that his results vary significantly around the average. A likely explanation for this difference is that the underlying quality differences among breakfast cereals over time are small, but quality differences for anti-cholesterol drugs are relatively large. This suggests that using demand estimates to construct quality-adjusted prices may be a promising approach for innovative markets where accounting for quality differences is particularly important. However, one should remain cautious in applying these techniques to consumer products where there is less innovation,

since quality-adjusted price indices may be sensitive to some basic assumptions. In addition to these checks, I also investigate how drug insurance may impact the price index. Although I find that drug insurance has a significant and positive effect on the demand for anti-cholesterol drugs, removing the effects of drug insurance has a relatively limited impact on the price index. This is a useful result as it suggests that commercial claims data may be used to construct similar price indices, even if the sample contains only insured individuals.

The next section provides a brief review of the literature. Section 3 describes the market for Cholesterol Drugs. Section 4 discusses the literature. Section 5 discusses the data, followed by a discussion of the results in section 6. Section 7 concludes.

2 A Brief Review of the Related Literature

There are several papers that have examined issues of price indices in drug markets, but it remains a challenge to incorporate new pharmaceuticals into a price index that accounts for changes in quality.[5] As mentioned previously, only the relatively recent and innovative work of Cleanthous (2004) and Lucarelli and Nicholson (2009) apply the discrete choice framework similar to Berry (1994) and BLP (1995) to analyzing new goods in pharmaceutical markets. A key advantage of this discrete choice approach is that it is relatively easy to estimate demand for a large number of products and to derive the value of newly introduced products.[6] This is important because new drugs often compete with many products in a class of drug treatments. Applying an approach similar to BLP that uses aggregate data and micro simulations, Cleanthous (2004) examines innovation in the market for depression drugs. He finds large welfare gains from the introduction of a new class of SSRI depression drugs. Lucarelli and Nicholson (2009) examine the demand for colorectal cancer drugs by estimating a demand model using aggregate data similar to Berry (1994). They find that quality-adjusted prices remain roughly constant over a 12 year period, although the unadjusted average price increased by hundreds of percentage points.

Both Cleanthous (2004) and Lucarelli and Nicholson (2009) examine the value of drugs using aggregate data. By contrast, in this paper I estimate the market demand of anti-cholesterol drugs using micro level data, which provides a number of substantial advantages. First, observing individuals and their reported diseases gives precise information on the risk factors driving the use of cholesterol-lowering drugs. If individual health information is not observed it may be difficult to separately identify a demand increase resulting from an improvement in the quality of a drug from one caused by the growing prevalence or awareness of a condition. Second, the micro data also allows for a more precise measurement of how individuals with a particular disease and demographic characteristics may demand treatment and it may be used to control for potential moral hazard effects caused by drug coverage (e.g. Which drug is preferred by an older individual with heart disease that lacks health insurance?). Many papers

[5] See Aizcorbe and Nestoriak (2010) for a more complete review of the price index literature for prescription drugs.

[6] An alternative to this approach is the multistage budget share methodology used by Fisher et al (1997) to analyze the demand for anti-ulcer drugs. This approach works well in some settings where there are a large number of products, but it is more difficult to incorporate consumer level information in this framework.

including Petrin (2003), Goolsbee and Petrin (2004), and Gaynor and Vogt (2003) have found the use of consumer level data to vastly improve estimates of demand. Finally, individual level information on health conditions is useful for the construction of disease-specific price indices.

In this paper the value of a new good is calculated based on the market's value of new technologies and how the market responds to changes in drug prices. This method contrasts with the production-based approach that measures the cost of health inputs relative to the production of health outcomes. One of the seminal papers examining the value of new medical technologies based on outcomes is Cutler et al (1998). They shows that new treatments for heart attacks increase the number of expected life years gained and that these health benefits exceed the cost of treatment. Cutler and McClellan (2001) look at the evidence across five conditions: heart attacks, low-birthweight infants, depression, cataracts, and breast cancer. They find that the benefits of technological change outweigh the cost in all conditions, except breast cancer where they find that the benefits and the costs are approximately equal in magnitude. Berndt et al (2002) find that the real cost of treating major depression decreased by about two percent per year between 1991 and 1996.

Although the two approaches are similar in their objective to measure the value of new technologies, they actually answer distinct questions that provide different insight into the value of new goods. The market-based approach is a reflection of the market's valuation of a product, while the production-based approach attempts to objectively measure the performance of the market based on cost-effectiveness studies that compare health outcomes and the cost of inputs. Numerous factors may cause these two approaches to lead to different valuations: (1) market distortions may cause the market valuation to be distinct from the cost-effectiveness studies (e.g. moral hazard, physician agency problems or asymmetric information); (2) the cost-effectiveness studies may be flawed (e.g. there may be factors that affect quality of life that may be difficult to quantify such as side effects or the hassle of taking a pill); (3) the market may have inaccurate information about the effectiveness of treatments (e.g. doctors may believe a treatment works and is cost effective, even though it is actually ineffective); (4) the researcher conducting cost effectiveness studies may have inaccurate information (e.g. the sample size may be too small or the length of the study too short); and (5) it is also possible that insufficient medical studies lead to ambiguity about the benefits of new technologies, leaving the market to price these uncertain benefits.[7] In the ideal case, absent market distortions, the two approaches should produce relatively similar results. However, identifying discrepancies between the two approaches may help identify market failures and policy solutions.[8]

This paper also relates to the growing health literature that measures the cost of disease treatment. Berndt et

[7] By ambiguity I am referring to the definition of Epstein and Zhang (2001) where they describe ambiguity as a state where there is insufficient information available for a decision maker to assign probabilities to events. This occurs in health care markets when new technologies are approved, although research about the effectiveness of technologies may continue throughout the life-cycle of a product.

[8] The production based approach may find results that are very different from the market's valuation of goods, and may impact the future market valuation. A simple example may be the drug Vioxx that had a high market demand, although it was later discovered that it had sufficiently high adverse side effects that it should be removed from the market. A carefully conducted production based analysis may assign little welfare to the widely used Vioxx treatment for arthritis prior to its withdrawal despite its popularity, while the value from the market based approach would be a reflection on the high demand for the product.

al (2001) advocate a "medical care expenditure price index" that tracks the cost by episode of an illness, which is a more suitable approach for analyzing medical care cost changes relative to a service price index that tracks the price of a particular service (e.g. a doctor office visit). Underlying this approach is the idea that individuals ultimately value the treatment of a condition and do not directly value the health inputs themselves. Progress in this area has generally followed the methodology laid out by Cutler et al (1998) that weigh the medical cost of treatments against the benefits. Although there have been some carefully conducted studies, as mentioned above, it is generally challenging to control for changes in quality of medical care inputs across a wide range of conditions. Moreover, even if one is able to control for the overall quality of an input price, typical price indices may not work well in this setting because the ability of an input to treat a condition and affect outcomes may depend on a patient's condition. The disease specific price indices presented here may be combined with disease expenditure estimates, such as those provided in Aizcorbe and Nestoriak (2010), to adjust the expenditures on anti-cholesterol medications to reflect changes in treatment quality for a specific condition.

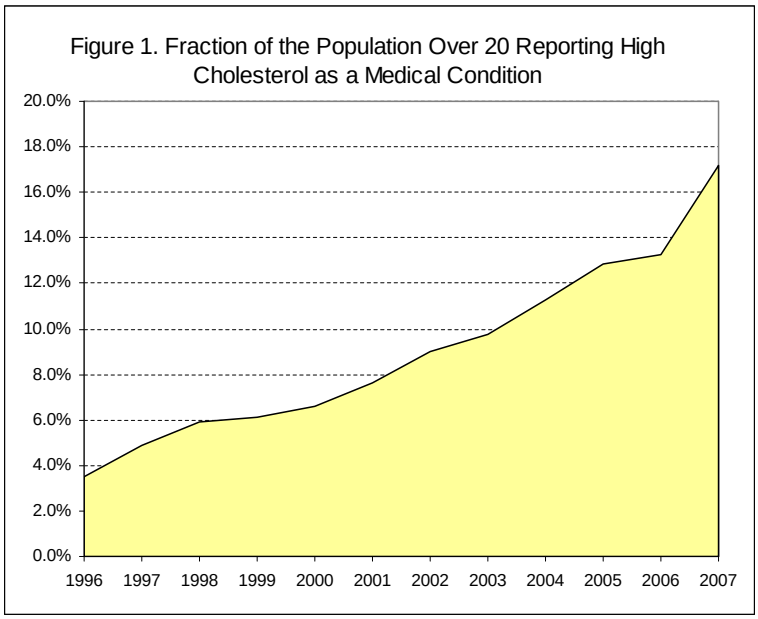
3 The Market for Cholesterol Drugs in the United States

For many patients the consequence of not taking cholesterol reducing drugs is detrimental to their health. According to the national treatment guidelines reported in National Cholesterol Education Program (NCEP) (2001), the primary goal of drug therapy for patients with high cholesterol is to attain lower LDL cholesterol levels. Evidence from epidemiological studies suggest that lower levels of LDL cholesterol (bad cholesterol) are associated with lower overall risk of clinical heart disease morbidity and mortality. As mentioned previously, high cholesterol is a contributing factor in 56 percent of clinical heart disease cases, but it is also a contributing factor for 18 percent of strokes. The World Health Organization (2002) reports that high cholesterol causes 4.4 million deaths in the world each year.

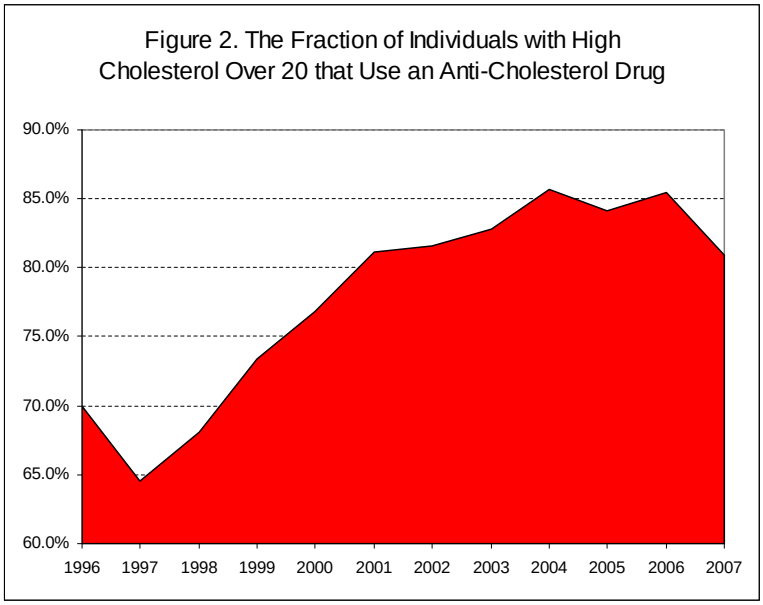
Significant improvements have been made in the treatment of high cholesterol. According to estimates from the Centers for Disease Control and Prevention, 28 percent of individuals over 20 had high cholesterol in the late 1970s prior to the introduction of the most effective anti-cholesterol drugs. That figure is around 16.3 percent today and much of the decline is likely attributable to the introduction of new cholesterol-lowering drugs and an increase in the number of individuals being treated.[9] The period of study in this paper, from 1996 to 2007, is a period of particularly rapid growth in both the awareness of high cholesterol and the use of anti-cholesterol medication. Figure 1 below shows the growth in the percentage of individuals in the U.S. 20 or older that report having high cholesterol.[10] The figure shows a five fold increase in the percentage of people reporting high cholesterol, from 3.5 percent in 1996 to 17.2 percent in 2007.

[9] These statistics are reported in Health United States (HUS) (2009). High cholesterol is defined as serum cholesterol levels of 240 or higher. The estimates are based on actual cholesterol readings, so those that would have high cholesterol absent medication are excluded from this calculation.

[10] These figures are from the MEPS data, discussed in greater detail in the data section. It includes individuals that would have high cholesterol without cholesterol lowering treatment.



In addition to a growth in the number of individuals reporting high cholesterol, there has also been an increase in the fraction of individuals with high cholesterol using anti-cholesterol medication. Figure 2 reports the fraction of individuals with high cholesterol over 20 that reported taking anti-cholesterol drugs. The growth in Figures 1 and 2 are likely caused by a number of factors.[11] First, over this period there has been mounting clinical evidence linking high cholesterol to heart disease and greater evidence of the effectiveness of cholesterol-lowering treatments (See NCEP (2001)). Second, more people may use anti-cholesterol drugs because more effective drugs have been introduced, but the growth may also be caused by the entry of low-priced generics. Third, there has been an increase in the level of advertising that may be related to the growing clinical evidence of drug effectiveness.



[11] The overall growth in the use of anti-cholesterol drugs is not unique to the selected data source. Similar findings of growth for cholesterol treatment are reported in Health United States (2009) that finds a ten fold increase in the use of statin drugs from the period 1988-1994 to 2003-2006.

Prescription drugs to treat high cholesterol have been around for more than four decades, but the introduction of new statin drugs in the 1980s have been revolutionary for the treatment of high cholesterol. The statin drugs have been proven to be the most effective at lowering LDL cholesterol, have few side effects, and are easy to administer.[12] This has led them to become the top selling class of drugs in the U.S. during the period between 1999 to 2008.[13] Compared to other cholesterol treatments, statin drugs are relatively new, with the introduction of the first drug Mevacor in 1987. Several drugs have entered the statin class since then including Pravachol, Zocor, Lescol, Baycol, Advicor, Vytorin, Lipitor and Crestor.[14] Table 1 below shows market shares of the various statin drugs from 1996 to 2007, along with the market share of non-statin medications. Arguably the most economically important entrant has been Lipitor, which entered the market in 1997 and became the top selling drug in the U.S. by 1999 and remained the top selling drug over the next decade.[15] At the time of Lipitor's entry into the market it was the most effective drug for lowering LDL cholesterol. Another important shift in cholesterol treatments has been the introduction of generic statins. This includes the generic version of Mevacor, which lost patent protection in 2002, and later generic versions of Pravachol and Zocor that lost patent protection in 2006. Prior to losing patent protection, Zocor was the second leading seller of anti-cholesterol medications, so it is not surprising that the generic versions of Zocor captured over 20 percent of the market by 2007.[16]

Table 1. Market Shares of Users of Cholesterol Drugs - MEPS Data

Drug Name	Chemical	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Lipitor	Atorvastatin Calcium	0.0%	11.8%	28.2%	34.6%	39.1%	44.3%	44.2%	45.2%	43.5%	42.0%	38.4%	32.2%
Zocor	Simvastatin	27.2%	28.1%	24.8%	25.6%	24.9%	26.2%	26.7%	25.1%	23.4%	21.7%	13.0%	4.4%
Generic Zocor	Simvastatin	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8.1%	21.5%
Pravachol	Pravastatin Sodium	21.8%	18.3%	17.1%	15.6%	12.8%	11.5%	11.5%	9.9%	8.2%	6.3%	3.4%	1.6%
Generic Pravachol	Pravastatin Sodium	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.2%	3.3%
Mevacor	Lovastatin	18.2%	12.6%	7.1%	5.1%	4.8%	2.7%	0.5%	1.3%	0.9%	0.3%	0.5%	0.6%
Generic Mevacor	Lovastatin	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.2%	4.3%	5.3%	8.1%	9.1%	9.3%
Crestor	Rosuvastatin Calcium	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	4.4%	4.6%	6.4%	6.9%
Baycol	Cerivastatin Sodium	0.0%	0.0%	1.3%	3.0%	4.7%	4.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Vytorin	Ezetimibe/Simvastatin	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	4.3%	7.6%	8.5%
Lescol	Fluvastatin Sodium	11.6%	12.1%	9.3%	5.7%	4.1%	3.5%	4.2%	3.6%	2.5%	2.2%	2.1%	1.2%
Advicor	Lovastatin/Niacin	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.4%	0.6%	0.6%	0.3%	0.3%
Other non-statins	Other non-statins	21.3%	17.1%	12.3%	10.5%	9.7%	7.2%	8.5%	9.8%	10.7%	9.9%	9.9%	10.2%

In addition to the statin class of drugs, there are four other classes of drugs that may be used to treat high cholesterol: nictonic acid derivatives, fibric acid derivatives, bile acid sequestrants, and ezitimbe. In general, these medications are less effective at reducing LDL cholesterol and have more severe side effects than the drugs in the

[12] Statins are typically the first drugs prescribed for the treatment of high cholesterol. They are an important part of treatment for the prevention of heart disease, stroke, atherosclerosis, and other atherosclerotic conditions. Atherosclerotic conditions include any condition related to the deposition of cholesterol that builds up as plaque on the innermost layer of the walls of large and medium-sized arteries. The active molecules in the statin class work by controlling the key enzyme that controls cholesterol in the body. The effect of statin drugs is that they lower LDL cholesterol (bad cholesterol) and triglycerides levels (also bad), and increase HDL levels (good cholesterol).

[13] "Statins Dethroned" Forbes (2009)

[14] Note that Baycol entered in 1998, had only 2% market share in its debut year. Its market share stayed small before it voluntarily withdrew in August of 2001 because it was linked to over 31 deaths caused by muscle cell damage.

[15] From IMS Health pharmaceutical sales estimates.

[16] Generic manufacturers can legally offer new products in a market using the active molecule of a drug when a branded drug's patent expires.

statin class; consequently the market share of these other drugs have declined from their 21 percent high in 1996 and have not exceeded 11 percent since 1998.[17]

Although the drugs in the statin class are generally more effective than other treatments, there are also key differences among statin drugs. Table A1 in the appendix displays attributes of anti-cholesterol drugs related to the effectiveness of each drug at lowering cholesterol. For example, it shows that Lipitor and Crestor being the most effectiveness at lowering LDL cholesterol, the primary target of drug therapy. There are many attributes not shown in Table A1. Drugs may also differ in their side-effects and proven effectiveness based on clinical outcomes. For instance, Zocor was one of the first drugs shown to be effective in clinical trials at reducing cardiovascular deaths.[18]

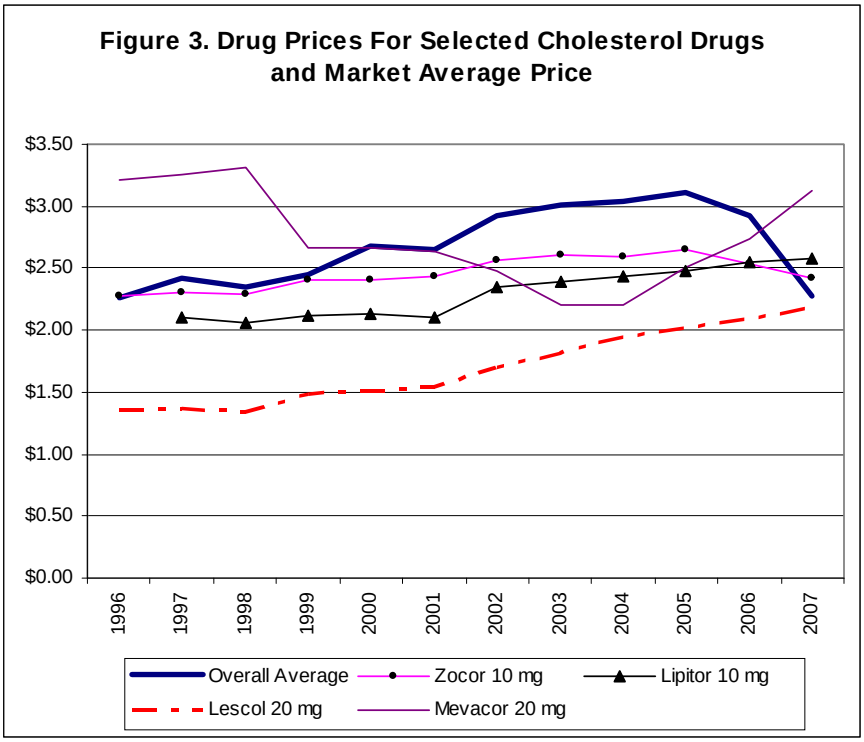
In addition to differences in effectiveness, anti-cholesterol drugs also differ in their pricing. Figure 3 shows the average price of a dose of treatment along with prices of select drug treatments.[19] The average price increases from 1996 to 2005 because of a growing demand for newly introduced drugs that tend to be more expensive and prices have trended upward on many of the more popular drugs (i.e. Lipitor, Zocor, and Pravachol). Each of these drugs was patented with a unique molecule and for much of the sample these patents were enforceable. As a result, generic firms could not enter the market, and prices remained relatively high. The typical cost per day ranges from about \$2 to \$3 for branded drugs. The generic drug prices tend to be much lower with the generic versions of Zocor and Pravachol costing 75 percent less than their branded versions in 2007.

[17] The second most effective class for lowering LDL cholesterol are bile acid sequestrants. While bile acid sequestrants are nearly as effective at lowering LDL levels as some of the statin drugs, bile acid sequestrants are used less often because of their severe side effects, which includes gastrointestinal and other problems. The other two classes of drugs, fibric acid derivatives and niacin, also have more severe side effects and are the least effective at lowering LDL cholesterol.

There are also some combination drugs that have entered the market, such as Vytorin and Advicor. These combination drugs include statin molecules but account for less than 9 percent market share in all years.

[18] See the Scandinavian Simvastatin Survival Study (1994).

[19] The overall average price is greater than those for the selected drugs because I have excluded many of the more expensive higher dose treatments.



Before proceeding to the demand model, note that Figures 2 and 3 provide conflicting descriptive evidence regarding welfare changes. If Figure 2 is viewed as a quantity index then one might infer, through revealed preference, that individuals are better off in 2005 than in 1996 because more individuals with high cholesterol are taking anti-cholesterol medications. However, looking at the increase in average price in Figure 3, one might conclude that welfare has declined. The next section presents a demand model that may be used to estimate welfare directly, which is used to measure the relative importance of the effects from the changing prices and shifts in quality.

4 Econometric Model of Demand

In contrast to most purchasing decisions, in prescription drug markets individuals rely on their doctors to tell them which drug, if any, is best suited to treat their condition. At the same time the insurer induces price sensitivity through the structure of the insurance plan, which is important since the full price of the selected drug ultimately has an effect on premiums. In the case where the doctor and insurer act in the best interest of the individual, the individual is able to optimally choose a medication. This is the maintained assumption throughout the presentation of the model. However, to the extent that market distortions are present, then the model below will only be an approximation to individual utility, and may be more appropriately viewed as a market demand function. In other words, if distortions are present, then the following choice model of the "individual" should be considered the joint decision of the individual, doctor, and insurer.

In each period I assume that individuals choose a product that maximizes their utility. The set of options is $\{0, ..., J_t\}$ where $J_t + 1$ is the number of products available in period t. Here the option 0 is the choice not to take

a drug. Each individual only chooses one option. Individual i chooses option $j \in \{0, ..., J_t + 1\}$ in period t if

$$u_{ijt} > u_{ikt} \quad \forall k \neq j.$$

I assume that individual i's indirect utility for product j where, $j \neq 0$, at time t is given by

$$u_{ijt} = \alpha_{it} p_{jt} + \beta_{it} x_{jt} + \xi_{jt} + \epsilon_{ijt},$$

where p_{jt} is the price of drug j in period t, x_{jt} is the vector of characteristics of drug j in period t, ξ_{jt} is a mean product specific error term, and ϵ_{ijt} is the idiosyncratic component of an individual's utility for drug j. The mean utility of the outside good is normalized to be zero. The response of individual i to the price and product characteristics consists of a component that is common to all individuals and a component that depends upon her observed characteristics, z_{it}:

$$\alpha_{it} = \alpha_0 + \alpha_1 z_{it}$$

and

$$\beta_{it} = \beta_0 + \beta_1 z_{it}.$$

Estimating Equations: To estimate the above model I follow the approach of Berry, Levinsohn, and Pakes (2004) who discuss estimation of the above model using micro data. Although it appears that the model can be estimated using a simple conditional-logit model, it is likely that the price variable will be endogenous. In fact, several studies have found evidence of price endogeneity, despite using micro level data, including Villas-Boas and Winer (1999), Gaynor and Vogt (2003), Goolsbee and Petrin (2004), and Chintagunta et al (2005).

The estimation procedure has two stages. In the first stage, it is helpful to note that the mean component of the utility of individuals choosing drug j at time t is a linear function of price, product characteristics, and the mean product specific error: $\delta_{jt} = \alpha_0 p_{jt} + \beta_0 x_{jt} + \xi_{jt}$. Also, note if one assumes that ϵ_{ijt} takes on an extreme value distribution then the probability of choosing option j takes the logit form:

$$Prob_{it}(j|z, x, \delta, \alpha, \beta) = \frac{exp(\delta_{jt} + \alpha_1(z_{it})p_{jt} + \beta_1(z_{it})x_{jt})}{\Sigma_{k=0}^{J_t} exp(\delta_{kt} + \alpha_1(z_{it})p_{kt} + \beta_1(z_{it})x_{kt})} \tag{1}$$

Therefore, in the first stage of the estimation, the equation 1 is estimated by maximum likelihood which identifies the α_1 and β_1 parameters of the model along with mean utility δ_{jt}. Note that when one has individual level data, then δ_{jt} may be estimated directly using maximum likelihood, so it is not necessary to solve for δ_{jt} as is typical when only aggregate level data is available. After the δ_{jt} parameters have been computed, they are used as regressors in the second stage estimation, where mean utility is regressed on price and other factors.

$$\delta_{jt} = \alpha_0 p_{jt} + \beta_0 x_{jt} + \xi_{jt} \tag{2}$$

When estimating the second stage, it is important to consider the possibility that the price variable is endogeneous. I address issues of endogeneity using both product-strength fixed effects and instrumental variables.

Instruments: It is often challenging to find valid instruments that affect a firm's pricing but are uncorrelated with unobserved quality measures, ξ_{jt}. Common instruments are factors that affect cost or rival product characteristics. However, the instruments used in this paper exploit the detailed micro-level data used in the estimation of the first stage of the model.

The basic intuition behind this instrumental variable (IV) technique is that an individual's choice is determined by her characteristics when selecting a product, as reflected in the demographics that enter the first-stage choice model. However, since individual information is conditioned out of the model in the first stage, it should not enter the unobserved component of demand, ξ_{jt}. Therefore, individual demographics should not be correlated with the unobserved component of demand; but the aggregate preferences of individuals in the market should be correlated with the price, because profit maximizing firms will consider the overall market demand (including population characteristics) when setting price. Kennan (1989) is the first paper that I am aware of that suggests that aggregate demographic information may be used as an instrument for price in the context of a linear model using micro data, although he presents no formal proof. Section 10.3 of the appendix provides a more detailed discussion regarding the use of aggregate demographics as instrumental variables in the context of a simple linear demand model.

The approach taken here is closest to that of Gaynor and Vogt (2003) who account for endogeneity when estimating a discrete choice model using micro level data.[20] Gaynor and Vogt demonstrate that the first-stage demand estimates from the logit maximum likelihood equation as an instrument for price. Similar to Gaynor and Vogt, I assume that firms choose price based on a mark-up term derived from an oligopoly pricing model that depends on both the demand for the product and the derivative of demand with respect to price. Both the demand function and the derivative may be calculated by summing individual decisions and their responses to price. Specifically the market demand for product j at time t is simply:

$$D_{jt} = \sum_{i=1}^{I} Prob_{it}(j|z, x, \delta, \alpha, \beta) \tag{3}$$

and the responsiveness to price is measured as:

$$\frac{\partial D_{jt}}{\partial p_{jt}} = \sum_{i=1}^{I} \frac{\partial Prob_{it}(j|z, x, \delta, \alpha, \beta)}{\partial p_{jt}}. \tag{4}$$

Using equations 3 and 4 one can show that the estimated mark-up is:

$$p_{jt} - mc_{jt} = \frac{\frac{\partial D_{jt}}{\partial p_{jt}}}{D_{jt}}.$$

[20] Another interesting example of demographics being applied as instruments is Romeo (2010) that uses consumer demographics as instruments in a discrete-choice model with random coefficients using aggregate data.

The first-stage estimates may be used to construct these demand measures, but they are likely to be endogenous because the function $Prob_{it}(j|z,x,\delta,\alpha,\beta)$ depends on the market price, p_{jt}, and the unobservable, ξ_{jt}. Therefore, in order to use the first-stage estimates, the terms containing price, p_{jt}, and the unobservable, ξ_{jt}, must be removed from the equation; so to construct the instruments all parameters interacted with price, α_{it}, are set equal to zero and I also set δ_{jt} equal to zero. That is, 3 and 4 are estimated at the point where $Prob_{it}(j|z,x,\delta=0,\alpha=0,\beta)$. With these terms set to zero, the remaining components of the estimated mark-up are used to construct the instruments $D^I_{jt}(j|z,x,\delta=0,\alpha=0,\beta)$ from 3 and $\frac{\frac{\partial D^I_{jt}(j|z,x,\delta=0,\alpha=0,\beta)}{\partial p_{jt}}}{D^I_{jt}}$ from 4. I allow the instrument for generic drugs to be distinct from the branded drugs. Since generics often compete with other generics and may also have costs that are different from the branded firm's, I construct a second set of instruments where I interact a generic dummy with the two instruments, $generic_{jt} \cdot D^I_{jt}$ and $generic_{jt} \cdot \frac{\frac{\partial D^I_{jt}}{\partial p_{jt}}}{D^I_{jt}}$.[21]

Aggregating over individual demand produces valid instruments in the above model; but it is instructive to observe that population characteristics may produce bias demand estimates if individual information did not enter the above model. Absent the inclusion of individual level data, any average change in utility caused by a change in the population's mean age or health condition would enter the error term, ξ_{jt}. Therefore, if demographics are used as an instrument in this setting it is likely that both ξ_{jt} and p_{jt} will be correlated with the instruments and the estimates would be bias. However, when micro data is included in the first stage of the estimation, then by construction, $\beta_1(z_{it})x_{jt}$, should be uncorrelated with the error term, ξ_{jt}.

Discussion. When presenting the model above, I assume there are no agency problems between the doctor and her patient. However, this assumption is not necessary for the purposes of this analysis. Irrespective of the interpretation of who makes the decision to purchase a prescription drug, the model presented here may be viewed as a model of demand that indicates the value the market places on anti-cholesterol drugs. In other words, even if market distortions are present, the above model still reflects the combined response of insurers, physicians, and patients to prices and drug characteristics.

The model does not explicitly address the issue of the quantity of medication consumed during each period. I assume that conditional on purchasing an anti-cholesterol drug in a period, individuals have inelastic demand and are, therefore, unresponsive to price. There are a number of reasons to hold quantity fixed within each period. First, the model already accounts for quantity because individuals are observed over three periods within a year, so I observe whether or not they purchase in each of the three periods. Quantity is also accounted for because I treat each strengths of a drug as a distinct product. Second, to the extent that compliance does not vary over time for individuals, this assumption will have little effect on the price index calculation. Finally, one appeal of the proposed approach is that it incorporates a large amount of individual data and is relatively simple to implement; addressing

[21] One complication with constructing the estimate for $\frac{\partial D^I_{jt}}{\partial p_{jt}}$ is that it depends on α, which is not observed. To address this problem I estimate an alternative demand model where I use D^I_{jt} and $D^I_{jt} \cdot generic_{jt}$ to instrument for price. I then use the estimate of α from this IV regression to obtain an estimate of $\frac{\partial D^I_{jt}}{\partial p_{jt}}$.

the issue of the quantity of medications purchased may unnecessarily complicate the model.[22] Within each period I assume that individuals are 75 percent compliant.[23]

4.1 Quality-Adjusted Price Measures

The quality-adjusted price index in this paper is based on the changes in the compensating variation derived from the estimated demand model. The compensating variation provides a measure of how much prices would need to change across the two periods to leave individuals indifferent between the old choice set and the new choice set. Given the logit functional form, the compensation variation from period $t-1$ to period t for individual i is calculated as

$$\Delta W_{it} = \frac{\mathrm{E}(u_{it}) - \mathrm{E}(u_{it-1})}{\alpha_{it}},$$

where $\mathrm{E}(u_{it})$ is the unconditional indirect utility and α_{it} is the marginal utility of income. The value of the unconditional indirect utility may be computed by integrating over the extreme value distribution. Using the derivation of McFadden (1981) the unconditional compensating variation is computed as

$$\Delta W_{it} = \frac{\sum_{j=0}^{J_t} \ln(\alpha_{it} p_{jt} + \beta_{it} x_{jt} + \xi_{jt}) - \sum_{j=0}^{J_{t-1}} \ln(\alpha_{it} p_{jt-1} + \beta_{it} x_{jt-1} + \xi_{jt-1})}{\alpha_{it}}$$

As described in greater detail by Trajetenberg (1990), the compensating variation can be converted into a price index by solving for the factor by which all prices are multiplied in period t in order to get the same welfare effect as ΔW_{it} for each individual. More precisely, given the change in welfare ΔW_{it}, to estimate the change in "real" price one solves for the value φ_{it} that solves the following problem:

$$\Delta W_{it} = \frac{\sum_{j=0}^{J_t} \ln(\alpha_{it} p_{jt} \cdot (1+\varphi_{it}) + \beta_{it} x_{jt} + \xi_{jt}) - \sum_{j=0}^{J_t} \ln(\alpha_{it} p_{jt} + \beta_{it} x_{jt} + \xi_{jt})}{\alpha_{it}}$$

If welfare increases across the two periods, then φ_{it} will be a negative value; and if welfare decreases across the two periods, then φ_{it} will be a positive value. The index will be specific to each individual in the data, and depend on her observed characteristics. Note that the above price index depends on current period prices and product characteristics which produces more conservative estimates of price changes relative to an alternative measure that uses base period prices and product characteristics. In fact, when the price index is computed from the base period

[22] One might also even view the level of compliance as a distinct decision model. While the doctor talks with individuals periodically about their condition and are involved in the overall decision to prescribe medication, they are not typically reminding individuals to take medication on a daily basis.

[23] The compliance rate observed in my data is around 74% on average (measured by number of pills/days in the round). Wosinska (2002) reports that out of a 43 day period, consumers typically miss 13 days of treatment, which is a compliance rate of around 70%. I choose a figure slightly higher than these because these compliance rates exclude free samples.

$(1 + \varphi_{it})$ may even be negative. In other words, for sufficiently large innovations individuals may prefer the new innovation more than having the base period products offered at a zero price.

To solve for the value of φ_{it} I apply a simple iterative search procedure for each individual. To construct an aggregate price index, I average across individual price changes. Similarly, to construct a disease-specific index, I aggregate over individuals with a specific health condition.[24]

There are some assumptions that should be considered when constructing quality-adjusted prices from demand estimates. Nevo (2003) shows that price indices derived from demand estimates may be impacted by whether one views unobserved demand, ξ_{jt}, and trend variables as representing changes in the "taste" for a product or changes in actual product attributes. In particular, one might be concerned that there is simply a growing trend in the treatment of high cholesterol that represents a growing "taste" for anti-cholesterol medications, although the products (and studies on the effectiveness of the products) have not changed. To address these issues I examine alternative estimates that fix the trend variable to its initial 1996 level and the error terms, ξ_{jt}, are also fixed over time. These two adjustments fix the market valuation of the different drugs over time.[25] The presence of drug insurance creates another concern. Drug insurance may cause a divergence between the private value of a product and its social value (i.e. a moral hazard effect). Cleanthous (2004) also addresses this point; in accordance with his approach, prior to calculating welfare, I remove the effect of drug insurance for all individuals in the data. The main result presented in this paper will remove the effects of drug insurance on demand, but allow the trend and product specific error to vary over time. However, I explore the effects of these assumptions by calculating price indices applying alternative assumptions.

Hedonic Price Index. I will contrast the quality-adjusted price index with two alternative price indices. One index is simply the average price in the market. The second index is constructed using a hedonic approach, which relies on measurable effects of these drugs on cholesterol levels. The anti-cholesterol drugs are a well suited category of medications to apply hedonic analysis because individuals primarily take these drugs to lower LDL cholesterol, which is a measurable attribute of all anti-cholesterol drugs (See Table A1 of the Appendix). The hedonic model is estimated by regressing the log of price on the characteristics of the drug, C_j, and time dummies, t. The hedonic regression is the following:

$$\log(p_{jt}) = \beta_c C_j + \eta_t + e_{jt}$$

Using the standard approach described in Aizcorbe and Nestoriak (2010), the hedonic price change from period t to period $t+1$ is $\frac{\log(n_{t+1})}{\log(n_t)}$. I also explore the hedonic approach advocated by Pakes (2003) that allows for greater flexibility in how product attributes relate to price.

[24] In constructing the aggregate price index, I weigh each individual by their population weights and the amount of welfare they receive from anti-cholesterol drugs. Whether individual weights are applied has little influence on the results. For instance, focusing on the median price change or an unweighted average produces similar results.

[25] However, as discussed later, I permit the molecule specific age variable to change over time to allow for a systematic increase in the value of newly introduced products because it typically takes time for the market to realize the full value of a newly introduced product.

5 Data

The main data source used in the demand estimation is the Medical Expenditure Panel Survey (MEPS) from 1996 to 2007. The survey contains extensive information on medical care in the United States, and it is used to provide national estimates on health care use, medical expenditures, and insurance coverage for the U.S. civilian, non-institutionalized population. The MEPS selects a random sample of households and surveys all individuals in a household. It follows the individuals for two years, during which it records information on individuals over 6 periods, where each period is approximately 4-6 months.[26] The data recorded in each period includes details on the individual's insurance, demographic characteristics, health condition, and medical expenditures. The MEPS study supplements the survey data by contacting the individual's medical providers and pharmacies to obtain billing information. For instance, if a patient reports purchasing Zocor from a specific pharmacy, the pharmacy is contacted to provide a payment history for all purchases of Zocor for that individual. Each year approximately 15,000 individuals enter the data so the data set is an overlapping panel.

The survey reports whether individuals have medical or prescription drug insurance and the type of plans. If the insurance plan is public, then the data identifies the insurance provider, that is, whether it is Medicare, Medicaid, or some other public agency.[27] Unfortunately, the MEPS data does not provide detailed information on the structure of the individual's drug insurance plan. The MEPS only contains information on payments for drugs purchased by an individual, but not on drugs that are not purchased.[28] For example, if an individual purchases Zocor, the data shows the out-of-pocket cost for Zocor, but not for the other anti-cholesterol drugs that could have been purchased.

Individuals are asked to write about their current medical condition and health history, including when their medical problems began. For each medical event (e.g., doctor visit or prescription drug purchase), individuals are asked about the medical conditions that gave rise to the event. Professional coders take the information provided by the individual and assign one of 5 digit ICD-9 codes (International Classification of Disease Code, Ninth Revision) which describe the individual's medical condition. To protect the identity of individuals in the sample, the 5 digit ICD-9 code is aggregated into 3 digit ICD-9 codes. The 5 digit ICD-9 codes are also aggregated into 260 clinically meaningfully categories using Clinical Classification Software. In this paper, both the 3 digit ICD-9 codes and clinical classification codes are used to describe an individual's medical condition. After reviewing risk factors mentioned in the ATP III report, the 3 digit ICD-9 and clinical classification codes are placed into four categories:

[26] While their are actually 5 rounds to the survey, the third round reaches across two years and is split into two distinct periods.

[27] Medicare provides medical insurance but no prescription drug insurance until the passage of part D in 2006, whereas Medicaid provides both. The data on private plans includes whether the plan covers doctors visits, prescription drugs, or other services. Additional information about the individual's insurance coverage can be inferred from the individual's medical expenditures. Each time a consumer visits a doctor or purchases medical services such as prescription drugs, the survey records the amount charged and who pays, whether the payment is paid directly by the consumer or paid by a third party. The third party payments are classified as private, Medicare, Medicaid or various other types of public insurances.

[28] There is a wide variety of features that an insurance plan might have such as formulary restrictions, deductibles, and copayments that may be fixed or vary across drugs.

cholesterol disorders, heart disease, diabetes and hypertension.[29] While most of the categories consist of only a single code, I define heart disease very broadly to include many severe conditions, such as stroke, heart attacks and other related conditions associated with the thickening or hardening of arteries. All of these problems listed are chronic conditions, so that once an individual is observed as having the condition, she is assumed to continue to have the condition.

The prescription drug transaction data provided in the MEPS includes the quantity, the strength, and the National Drug Code (NDC) of each drug purchased. The NDC code is a number that uniquely identifies a drug and can be used to link the drug to the manufacturer and a specific product. Conversion from the NDC code to a specific product is done using Redbook data that links NDC codes to the products and manufacturer.[30] In cases where the NDC code of the drug is not listed, I used the name of the drugs active molecule as listed by the pharmacist, and whether the drug is indicated as branded or generic to determine the identity of the anti-cholesterol drug.

The MEPS data includes price information, but appears to contain imputed prices based on average wholesale prices, especially in the later years of the data. This is problematic since average wholesale prices are typically much higher than the actual price of the drug. As an alternative, I estimate price information using MarketScan claims data to impute the price of prescription drugs for the period 1999 to 2007. The MarketScan data is a convenience sample from commercial insurers and large employers that is not representative, but is much larger, including several million individuals in most years. In addition to the large sample size, the key advantage of the MarketScan claims data is that it is based on adjudicated claims that contains more precise price information and does not contain imputations.

For the analysis that follows, I limit the sample to those with either a cholesterol disorder or heart disease. Based on this selection rule, the total number of individuals included in the analysis is 21,991 and the number of individual periods is 106,510. The fraction of individuals using anti-cholesterol medication that were excluded from the sample is 0.48 percent. The percentage of users of anti-cholesterol medication outside the selected sample likely represents people with other types or combinations of risk factors such as diabetes, hypertension, or a family history of heart disease who may also be taking statin drugs.[31]

[29] The three digit ICD9 and clinical classification codes used for each disease category are the following: Cholesterol Disorder (ICD9: 272), Diabetes (ICD9: 250), Hypertension (ICD9: 401), and Heart Disease (ICD9: 410-414, 433-437, 440, 444, and classification codes: 101, 104, 108, 109, 113, 114, and 116). Dr. Karen Rasmussen assisted in the assignment of these categories.

[30] One case where a unique drug cannot be assigned to a round is when multiple anti-cholesterol medications are purchased. Since multiple medications are typically not prescribed, it is likely that a patient has switched drugs. So in cases where two drugs are purchased in a round, I assign the last drug taken in the round. I use information on the drug taken in the previous or following round to determine the drug that a person is switching to. If that information is not available I assign the drug with the greatest quantity purchased in the round.

[31] I performed three checks on the MEPS data to determine how whether the sample is representative. I compared its estimate of the number of uninsured to that reported in the Census for 2002 and found that they matched. The Census estimate of the number of uninsured is 45.8 million, while the number from the MEPS is 43.6 million. I also computed the annual estimated national revenue shares in the MEPS of the top three sellers - Lipitor, Pravachol and Zocor - and compared them to the those reported in IMS health, which is a pharmaceutical market research firm that monitors drug sales from pharmacies. It reports total revenue data for the Statin class and

5.1 Variables

The dependant variable used in this paper is the treatment choice in a period. The treatment choices include the anti-cholesterol drugs that are available in the market in various strengths during the period and the no-drug treatment option. The dependent variable is a binary variable that is equal to 1 if individual i uses drug j in period t. I assume that if the individual takes any medication in a period, then she is considered to be using medication in that period.

I turn next to a description of the variables of the model. Individual i's health conditions in period t is described by four dummy variables: $High\,Cholesterol_{it}$, $Heart\,Disease_{it}$, $Diabetes_{it}$, and $Hypertension_{it}$.[32] Since cholesterol levels tend to increase with age, and men are at a higher risk of heart disease at a younger age, I also include the variable Age_{it} and an indicator for $Male_{it}$ and nonlinear functions of these variables. In addition to these objective risk factors, I also observe a subjective risk measure where individuals indicate their perceived health. The variable $PerceivedGoodHealth_{it}$ is an indicator that is one if an individual perceives there health as excellent, and zero otherwise.

The various health related variables mentioned in the previous paragraph are used to construct a measure of composite risk which I call $RiskScore_{it}$. While an ideal risk measure would be computed by weighting the risk factors based on likely health outcomes, this information is not available. Instead the composite risk is constructed by estimating a probit model of whether individuals in the sample take an anti-cholesterol drug conditional on the above risk factors, and then setting $RiskScore_{it}$ to be the predicted probability based only on health factors. The probit estimation used to construct $RiskScore_{it}$ is shown in Table A2 of the appendix.

I use binary variables for insurance coverage. The variable $DrugIns_{it}$ is equal to 1 if the individual i has drug insurance in period t and zero otherwise. An individual is classified as having prescription drug coverage if she has a private prescription drug insurance (including Medicare Part D) or is on Medicaid. This definition of drug coverage should account for nearly all individuals with drug insurance.[33] In this data, I find that those with drug insurance

for each of the three top sellers for the years 1999-2002. I found the differences between the samples to be relatively small. Finally, I compared the market shares from the MEPS to those from the MarketScan data and found them to have similar trends in market shares, despite having different samples.

[32] There was one change in the survey in 2007 that had a noticeable impact on the reporting of chronic conditions like high cholesterol, heart disease, hypertension, and diabetes. Prior to 2007 the MEPS survey asked individuals to list their conditions, but in 2007 the survey was changed by specifically asking individuals if a physician told them whether they had any of these chronic conditions. That is, they were specifically asked whether a doctor has told them that they have high cholesterol. It appears that this change may have had a measurable impact on the reporting, with those reporting a cholesterol condition jumping 4 percent to 17.2 percent of the population. This effect may also be seen in Figures 2 and 3 as a slight increase in the fraction of individuals reporting high cholesterol and a slight decrease in the fraction of individuals using anti-cholesterol drugs. Although this change does not directly impact the model, it will require a slight modification in the construction of the quality-adjusted price index for 2007, which is discussed in greater detail in the section on quality-adjusted prices.

[33] To account for the possibility of misreporting, I use prescription drug expenditure information provided by the MEPS to mark individuals as covered if a third party pays for a significant amount of their drug coverage for the year. I broaden the definition of those with prescription drug insurance by counting individuals as insured if 70% of their drug expenditures are covered by another party.

pay 28.9 percent of prescription drug expenses out-of-pocket.

The variable $MedIns_{it}$ is equal to 1 if individual i has medical insurance in period t and zero otherwise. Medical insurance coverage typically covers doctor office visits and other services, which makes it more likely that individuals will visit a doctor to obtain a prescription, even when they do not have drug insurance. Individuals on private plans, Medicaid, Medicare, or other public insurance plans are classified as medically insured.[34] Dummy variables are included to indicate whether and individual has either $Medicare_{it}$ or $Medicaid_{it}$ insurance. The model also includes information on individual i's household income in period t and is measured in 2007 dollars as $Log(Inc_{it}+1)$. It also includes the number of years of education, $EducYear_{it}$.

The characteristics of the drugs that are invariant over time are captured using drug–strength specific dummies. The drugs in the market are the statin drugs: Lipitor, Baycol, Pravachol (generic and branded), Zocor (generic and branded), Lescol, Mevacor (generic and branded) Crestor and the non-statins listed in Table A1 of the appendix. Many of the drugs are offered in multiple strength so that different strength categories are considered distinct products.[35]

The value of anti-cholesterol drugs may systematically vary over time. Given the large expansion in the use of anti-cholesterol drugs over the studied period, a trend variable, $Trend_t$, is included in the model to capture general shifts in treatment practices.[36] In addition to the trend variable, the model includes a variable that changes with the age of a newly introduced molecule, $log(AgeMolecule_{jt})$, to account for the fact that it may take time for the market to realize the value of a new molecule.[37]

The price of drug j in period t is denoted $Price_{jt}$. The price of the drug is the full price of the drug paid to the retail pharmacy (i.e. the amount paid by the insurer plus the amount paid out-of-pocket by the individual). In addition to not observing the co-payments for all available drugs, the total payment is used because the goal of the model is to measure the total market value of the product, and as Cutler et al (1998) argue, individuals ultimately bear the full cost of the payment through higher out-of-pocket costs, higher individual premiums, or lower wages (for employer paid premiums). Even if I observed the co-pays for the different treatment options, this would not necessarily capture the market's response to the full price of the prescription drug. In particular, it may ignore the price sensitivity of individuals as reflected in their selection of insurance options. For example, a person who is both highly risk averse and highly price sensitive, might prefer a plan that covers the full price of the lowest cost drug option, but provides no coverage for alternative drug choices. This person would have no elasticity based on

[34] I also assume that individuals with prescription drug insurance coverage also have medical coverage because it is rare for individuals with drug insurance not to have medical insurance

[35] The less frequently used strength categories are aggregated with the more frequently used strengths that are closest in value. For example, the 5 mg strength category for Zocor is purchased infrequently, so it is aggregated with the 10 mg category. Appendix A1 provides a list of the different categories used in the estimation. I found that the results presented here are not sensitive to alternative aggregations.

[36] The trend variable is the beginning date of the observations measured in years (e.g. 1997.82).

[37] The age of the molecule is the median date in the current round minus the date in which the molecule was approved for sale by the FDA divided by 365. I assume the effect of the molecule's age stops after 10 years, so the maximum value of this variable is log(10). The results are robust to alternative assumptions, such as not setting a limit on the age variable.

observed copayments, although they may be highly sensitive to market price. It is also important to note that even though drug insurance lowers the out-of-pocket payments, it is unclear how insurance affects the responsiveness to market price. The insurer may lower the out-of-pocket cost by a large amount, but still induce price sensitivity through tiering or formulary restrictions. I allow flexibility in how drug insurance affects the responsiveness to market price by including an interaction of $DrugIns_{it}$ with $Price_{jt}$, but I also allow drug insurance to have an effect on the probability of choosing any anti-cholesterol medication regardless of the price.

The price variable is calculated on an annual basis from transactions involving drug j. The task is complicated by the fact that what is observed in the data are transaction prices that vary by the strength of the drug per tablet, which is measured in milligrams, and the size of the bottle, which is measured in number of tablets. For example, Lipitor is available in strengths of 10mg, 20mg, 40mg and 80mg per tablet, and bottle sizes are typically 30, 60 or 90 tablets. Therefore, the number of prices for Lipitor is the number of strengths available times the number of bottle sizes, which in this case, is 12. In order to compare prices across different drugs and strengths, I choose a single price for each drug-strength category. To calculate price, for each drug-strength combination I run a regression that includes the different quantities of the drug purchased along with year dummies. The price of the drug-strength combination j in period t is the predicted value from these regressions for a bottle containing 30 tablets (the most frequently purchased quantity).

The identical estimation procedure was conducted on the MEPS data and MarketScan claims data for estimating price. As mentioned previously, I found a systematic bias in the later years of the MEPS data that appears to be related to how they impute prices.[38] Therefore, for the period from 1999 to 2007 I use estimated prices from MarketScan. The prices from the computations for a selected set of drugs and strengths is reported in Figure 3, presented previously. To match the time frame of the price variables, I estimate the mean utility, δ_{jt}, on an annual basis. An additional reason to estimate these variables on an annual basis is that many insurance plans have open enrollment once a year.

Interactions. To allow for flexibility in how individuals respond to the different prescription drug offerings the model contains a number of interactions. First, the model allows for several variables to affect price sensitivity through interactions with $Price_{jt}$, including the $RiskScore_{it}$ and $Log(Inc_{it}+1)$. Those with more serious health conditions may be less sensitive to price as are those with higher incomes. As mentioned previously, I also including

[38] I found that in the later years of the MEPS data, that the difference between the price of the branded and generic version of a drug were minimal, although the price difference reported by MarketScan were relatively large. One potential reason for the difference may be that the MEPS uses the average wholesale price of the drug to identify outliers (outliers are defined as those cases where prices are more than 20% below the average wholsale price), but in many cases drug prices are much different than the average wholesale price, especially for generic drugs. For instance, using the MarketScan data I find that the average wholesale price for 10 mg of generic Zocor in 2007 is \$2.70, but the average transaction price in the MarketScan data is \$1.81. So Zocor prices reported in the MEPS survey that fall close the the average price reported in Market Scan, would be identified as an outlier in the MEPS data and then replaced with the average wholesale price, which would introduce a very large upward bias on price. As a robustness check the model has also been estimated excluding the years that incorporate MEPS pricing data from 1996-1998 with qualitatively similar results to those reported for the full sample.

an interaction with price and the $DrugIns_{it}$ variable.

Interactions are included between the age of the individual and the age of the drug, to capture the possibility that older individuals and their doctors may be more familiar with prescription drugs that have been in the market longer. The model also allows for the affect of the severity of the patient's condition, as measured by $RiskScore_{it}$, to be interacted with the age of the drug and the trend variable. The first interactions allows for differing values of a specific molecule as they are introduced in the market, depending on the severity of the condition. The interaction with the trend variable allows for changing guidelines for cholesterol treatment over time. In particular, studies over this time period have suggested that individuals may benefit from more aggressive treatment, so that later in the period lower risk individuals may be more likely to purchase anti-cholesterol medications.

In addition to the above interactions, the regression contains interactions between three major risk factors (having high cholesterol, heart disease and age) and dummy variables for the active molecules for each of the anti-cholesterol drugs. The regression also includes interactions of these risk factors with an indicator of whether the drug is a generic. The ingredient left out of the interaction is the active ingredient for Zocor (i.e. Simvastatin), which is one of the more popular molecules and is in the sample for the entire period.

5.2 Summary Statistics

Table 2 provides descriptive statistics on the population in the selected sample. The first column of the table provides the mean of each variable, while the following columns show the quartiles. Overall the table shows considerable variation in many of the demographic variables. The distribution suggests that it may be challenging to model demand using more aggregate data (or applying random coefficients from the population) because the selected sample is quite distinct from the national population. The median age is 63 which is much higher than the national median age of about 35. This is not surprising since cholesterol increases with age as does the incidence of heart disease. A high fraction of individuals are enrolled in Medicare, so just 4 percent of the selected sample has no medical insurance, relative to the national average of about 16 percent. Table 2 also shows the prevalence of both hypertension and diabetes that are relatively more common in the sample compared to overall population.

Table 2. Demographics

Variable	Mean	25th Percentile	50th Percentile	75th percentile
Health Related Demographics				
Age	62.08	52	63	74
Health Index	0.47	0.20	0.58	0.67
Male	0.48			
Has High Cholestorol	0.68			
Has Heart Disease	0.48			
Has Diabetes	0.25			
Has Hypertension	0.53			
Perceived Health is Good	0.10			
Other Demographics				
Family Income (in 2007 $s)	$54,591	$18,663	$40,123	$74,755
Number of Years of Education	11.87	10	12	14
Drug Insurance	0.74			
Health Insurance	0.96			
Medicare	0.52			
Medicaid	0.16			
# of Observations	106,510			

6 Results

There are two key estimates to discuss: (1) the discrete choice demand model where the micro individual's logit choice problem is estimated and (2) the estimation of the components of the mean utility which separates mean utility into price and other factors.

Logit Demand. The first step of the estimation procedure is to estimate a discrete choice model to obtain estimates of mean utility and measure the impact of individual characteristics on the drug choice. Table 3 shows some of the key estimates from the first stage regression. The model shows that several factors affect price sensitivity which highlights a vertical dimension of product differentiation. Those with more severe conditions, those with drug insurance, and those with higher incomes tend to be less sensitive to price.

The estimates show that many of the risk factors are important determinants of whether an individual selects an anti-cholesterol medication with all of the risk factors having significant effects (i.e. the composite risk score, age, male, high cholesterol, heart disease, diabetes, perceived health, and hypertension). Note that even though the $Trend$ variable is increasing, indicating greater demand for anti-cholesterol drugs over time, the interaction between the $RiskScore$ variable and $Trend$ variable is negative, indicating that those with less severe conditions are more likely to take anti-cholesterol drugs later in the sample.[39]

The $\log(Age\ of\ Molecule)$ is another important determinant of the demand for anti-cholesterol medications. The estimates show a very heterogeneous effect on the age of the molecule depending on the characteristics of individuals, indicating how quickly new products are adopted by different segments of the population. The estimates show that

[39] The inclusion of the $RiskScore$ variable is a useful aggregate measure of an individual's health condition, but it complicates the interpretation of some of the other health variables, such as $HeartDisease$, because $HeartDisease$ is included as a part of the $RiskScore$.

older individuals are less likely to adopt new medications in favor of medications that have been in the market longer, perhaps due to greater familiarity with products that have been in the market longer. In contrast, individuals with higher risk conditions, as reflected by their risk score, are more likely to adopt new medications earlier, perhaps because they are more willing to risk potential problems from new treatments given the potential benefits.

Additional interaction terms not shown in Table 3 are included in Table A3 of the appendix which presents estimated parameters from the interaction of age, heart disease, and high cholesterol with the associated molecule of each drug and whether the drug is a generic. These interactions indicate that relative to other treatments, it appears that Zocor is preferred by patients for the treatment of heart disease. Also, younger individuals prefer both Lipitor and Crestor, which are the most effective medications for lowering LDL cholesterol on average.

Table 3. First-Stage Results from Conditional Logit Estimation

Variable	Coef.	z-stat
Price*Risk Score	0.218	(3.76)
Price*Drug Insurance	0.040	(2.34)
Price*Income	0.012	(1.76)
Drug Insurance	0.214	(3.88)
Health Insurance	0.490	(7.41)
Log(Household Income/1000+1)	0.003	(0.12)
High Cholesterol	4.308	(9.95)
Heart Disease	0.805	(10.85)
Age	0.118	(3.38)
Age^2	0.001	(1.73)
Age^3	0.000	(-5.07)
Age>=40	0.411	(4.14)
Age*Male	-0.011	(-5.4)
Perceived Good Health	-0.304	(-7.56)
Risk Score	2.106	(1.78)
Education	0.020	(4.96)
Medicare Health Insurance	0.052	(1.32)
Medicaid Health Insurance	-0.069	(-1.79)
Male	1.104	(7.11)
Hypertension	0.456	(9.74)
Diabetes	0.463	(9.2)
Log(Age Molecule)	-0.662	(-7.3)
Age*log(Age Molecule)	0.012	(8.04)
Risk Score*log(Age Molecule)	-0.174	(-1.62)
Trend	0.129	(5.87)
Risk Score*Trend	-0.160	(-7.54)
Number of Observations	106,510	
Psuedo R-Squared	0.444	

Reported Z-statistics are based on robust standard errors clustered by individuals.

Mean Utility Estimation. After estimating the logit model, the next step in the analysis is estimating the various components of mean utility. Recall that this is the second stage estimation from equation 2. Here I regress mean utility, δ_{jt}, obtained in the first stage on price and the exogenous variables. The exogenous variables are the prescription drug strength dummy variables, with the 10 mg version of Lipitor as the left out alternative. Table 4 reports the results from the two models. The first model is an IV estimation that accounts for the potential endogeneity of price. The results from the IV model show that the coefficient on price is negative and highly

significant. It is interesting to contrast the magnitude of the price coefficient, -1.81, with the coefficient on the interaction of price and drug insurance of 0.04 reported in Table 3. The relative magnitude of this interaction term implies that their is very little effect of prescription drug insurance on sensitivity to market price. This suggests thateven those with insurance are, in fact, quite responsive to the market price; although the presence of drug insurance does have a measurable and positive impact on the probability of taking prescription medication. To check the economic importance of applying the IV approach, the second column shows estimates from an OLS regression. The OLS model shows that the price coefficient is negative, but very small and insignificant; most likely this is due to a bias from firms that are able to charge higher prices on drugs with larger unobserved demand shocks.

Table 4. Estimates of Mean Utility on Price

	IV Estimation		OLS	
Variable	Coef.	Z-stat	Coef.	Z-stat
Price	-1.807	(-4.33)	-0.130	(-0.8)
Lipitor 20mg	1.096	(1.65)	-0.791	(-1.84)
Lipitor 40mg	0.727	(0.98)	-1.579	(-3.53)
Baycol .3mg	-3.801	(-4.94)	-2.650	(-4.4)
Baycol .4mg	-3.860	(-4.36)	-2.773	(-3.92)
Cholestrimine	-5.736	(-8.25)	-3.636	(-8.44)
Vytorin 20mg	-1.752	(-2.59)	-2.622	(-4.88)
Vytorin 40mg	-1.676	(-2.48)	-2.534	(-4.72)
Zetia	-2.339	(-3.85)	-2.889	(-5.86)
Fenofibrate	-1.120	(-2.28)	-1.552	(-3.89)
Lescol 20mg	-4.322	(-8.26)	-3.307	(-8.44)
Lescol 40mg	-3.811	(-7.38)	-2.852	(-7.3)
Generic Lopid	-3.884	(-4.24)	-0.696	(-1.42)
Lopid	-4.010	(-7.62)	-2.967	(-7.55)
Advicor	-3.062	(-5.5)	-3.035	(-6.58)
Generic Mevacor 20mg	-4.764	(-5.63)	-2.207	(-4.21)
Generic Mevacor 40mg	-4.356	(-6.14)	-2.589	(-5.26)
Mevacor 20mg	-1.958	(-3.96)	-2.707	(-7.01)
Mevacor 40mg	-0.941	(-1.23)	-3.400	(-7.58)
Generic Niaspan	-7.295	(-7.26)	-3.704	(-7.18)
Niaspan	-5.653	(-8.15)	-3.597	(-8.25)
Generic Pravachol 20mg	-6.466	(-6.43)	-4.271	(-5.85)
Generic Pravachol 40mg	-5.434	(-5.63)	-3.555	(-4.92)
Pravachol 20mg	-1.610	(-3.51)	-1.682	(-4.43)
Pravachol 40mg	-0.568	(-1)	-1.902	(-4.74)
Crestor 10mg	-1.167	(-1.88)	-1.944	(-3.92)
Crestor 20mg	-2.197	(-3.53)	-2.964	(-5.98)
Generic Zocor 10mg	-5.751	(-5.99)	-3.904	(-5.41)
Generic Zocor 20mg	-3.149	(-3.68)	-2.554	(-3.64)
Generic Zocor 40mg	-2.427	(-2.84)	-1.845	(-2.63)
Zocor 10mg	-1.503	(-3.25)	-1.764	(-4.64)
Zocor 20mg	2.105	(2.23)	-1.209	(-2.43)
Zocor 40mg	1.538	(1.73)	-1.520	(-3.15)
Constant	8.896	(8.75)	5.037	(10.84)
Number of Observations	266		266	
R-squared	0.381		0.575	

A number of checks are conducted to investigate the validity of the above model. Table A4 in the appendix shows the first stage regression results. The instruments used in this model have good explanatory power in the first stage. Applying a Cragg-Donald test for weak instruments, I find that the null hypothesis that the instruments are weak is strongly rejected. The model also produces reasonable price elasticity with a mean of -3.54 (s.d. 1.8) that

is consistent with profit maximizing behavior of drug manufacturers.[40]

A potentially important variable that is omitted in the above analysis is advertising to physicians. While this paper is not focused on the effect of advertising, one may be concerned that its exclusion may cause an omitted variable bias. To determine if this is a potential problem, I estimate an alternative model that uses advertising information available in the MEPS that indicates whether an individual received a free drug sample in a period. (Free drug samples are often provided to physicians to give to their patients and this is an important marketing tool.) Using the MEPS data I construct an aggregate measure of free samples given for a particular drug in a year. Including this variable of advertising in the IV regression is highly significant; but the price coefficient also remains highly significant and similar in magnitude, implying that the inclusion of the advertising variable does not affect the main result of the paper. Since the advertising variable is potentially endogenous, I focus on the analysis that excludes advertising information. Also note that even if advertising were in the model, it is unclear how it should enter the welfare analysis. Similar to the issue that arises with the error term, ξ_{jt}, the effects of advertising could represent an effect on individual taste or it may be informative and change the objective value of the product.

Welfare Analysis. Using the estimates above, the welfare for anti-cholesterol drugs is calculated in each year and reported in Table 5. The welfare growth is enormous, increasing from \$1.3 billion in 1996 to more than \$8.1 billion in 2007, a more than a 500 percent increase. Much of this growth in welfare is caused by an increase in the number of users, from 5.4 million in 1996 to 29.3 million in 2007. However, the growth is also partly due to an increase in the welfare per user of the drug, which has increased from \$241 per user in 1996 to \$278 per user in 2007.

Table 5. Welfare Estimates

YEAR	Total Welfare in Billions Per Year	Total Spending in Billions Per Year	Total Users in Millions	Welfare Per User Per Year
1996	\$1.31	\$3.34	5.42	\$240.74
1997	\$1.60	\$4.58	6.72	\$238.42
1998	\$1.85	\$4.95	7.86	\$235.18
1999	\$2.48	\$6.44	9.51	\$260.60
2000	\$2.92	\$8.38	11.51	\$253.39
2001	\$3.79	\$9.98	13.80	\$274.66
2002	\$4.24	\$12.75	15.95	\$266.15
2003	\$5.06	\$14.93	18.04	\$280.45
2004	\$6.10	\$18.29	21.79	\$279.82
2005	\$6.36	\$19.23	22.77	\$279.37
2006	\$7.01	\$19.40	24.19	\$289.64
2007	\$8.13	\$18.08	29.29	\$277.68

[40] In addition to these checks, I also estimate the model using alternative instruments. One set of instruments includes the number of competitors, the number of drugs that are generic, the number of generic competitors, and the number of competitors interacted with the age of the molecule. These variables are interacted with a dummy variable indicating whether the drug is generic. Another set of instruments includes only the estimated demand of the drug and the interaction of the estimated demand and whether the drug is a generic. Each of these alternative sets of instruments produce similar results to those presented here.

Table 5 shows aggregate welfare estimates, but individual welfare from the availabity of anti-cholesterol drugs is quite heterogenous. Table 6 shows expected welfare for individuals with different types of conditions for 2006.[41] The first row shows the welfare distribution for the entire population. The mean expected welfare per year is $193, but there is a wide range in welfare from anti-cholesterol drugs with the 10th percentile valuing the drugs at $110 and the 90th percentile valuing the drugs at $264. In general, Table 6 shows that those with more serious risk factors (i.e. heart disease, diabetes, hypertension, and age over 55) tend to value these drugs more. Financial factors also have a large impact on the value of these drugs. Both those with drug insurance and with health insurance value anti-cholesterol drugs more than those without insurance. The average difference in valuation for someone with health insurance compared to someone without health insurance is around $73 (=$199-$126), about the same effect as a serious risk factor, such as hypertension or heart disease.[42] Those with drug insurance also value cholesterol drugs more than those without drug insurance. Overall, Table 6 underscores the importance of controlling for heterogenous factors that influence individual valuations for these drugs.

Table 6. Welfare Per Year of Anti-Cholesterol Drug Availability by Condition and Demographic Factors

	Mean	Median	10th Percentile	90th Percentile
Overall	$193.09	$198.84	$109.78	$263.61
Heart Disease				
No	$179.94	$188.53	$98.87	$241.15
Yes	$240.93	$242.38	$178.04	$301.47
Has Diabetes				
No	$181.43	$188.12	$97.75	$244.72
Yes	$222.29	$226.55	$143.50	$292.02
Has Hypertension				
No	$159.91	$170.02	$73.96	$223.07
Yes	$215.33	$218.96	$154.03	$278.69
Age Greater Than 55				
No	$138.12	$146.56	$59.80	$201.97
Yes	$218.64	$217.64	$166.56	$276.43
Has Health Insurance				
No	$125.66	$109.78	$48.41	$230.09
Yes	$198.56	$201.96	$130.34	$264.88
Has Drug Insurance				
No	$160.20	$167.94	$78.06	$227.77
Yes	$202.97	$208.83	$130.34	$270.82

This section has used the demand estimates to look at welfare levels in the market, the next section examines how welfare changes over time may be translated into quality-adjusted prices.

[41] To reduce the level of heterogeneity caused by changes in prices or product attributes, I only select those individuals that are reported as having a cholesterol problem for the year 2006.

[42] This can also be seen by examining the coefficients in Table 3 above, which shows the effects of health insurance are comparable in magnitude to the effects of hypertension or diabetes.

7 Quality-Adjusted Prices

The demand estimates above may be used to construct a quality-adjusted price index. For comparison, the quality-adjusted index is benchmarked against two alternative price indices: one price index is a simple average price weighted by the number of users and the second index is a hedonic price index that is a frequency weighted regression of price on product characteristics and time dummies. Before showing the results of the different price indices, I first present the results from the hedonic regression in Table 7.

Three drug effectiveness measures are included in the hedonic regression: the medication's average effectiveness in lowering LDL cholesterol (bad cholesterol), effectiveness in increasing HDL cholesterol (good cholesterol), and the ability to lower Triglyceride levels (also bad). The regression also includes a dummy variable for whether the drug is a statin drug to capture the fact that statins are viewed as having fewer side effects. I find that only the LDL effectiveness is important in pricing anti-cholesterol drugs, which is consistent with the clinical guidelines that suggest the primary goal of drug therapy is to lower LDL cholesterol. The second model is identical to the first, but allows for a nonlinear relationship between the price and the LDL level. The second model shows that not only are prices higher on drugs that are more effective at lowering LDL cholesterol, but that each additional unit of effectiveness has a greater impact on price. Since the nonlinear effect on LDL levels appears to be important and the fit of this model is better than the linear model, a price index is constructed using the second model.[43] As an alternative to the hedonic price index presented in Table 7, another hedonic index was constructed following the methodology of Pakes (2003) by letting LDL cholesterol have a different relationship with $\log(price)$ in each year. Those results from the Pakes model are similar to those of Model 1 and Model 2, although less stable, perhaps because of the relatively small sample size.[44]

[43] Both models weight observations by the number of users of each type of drug. The unweighted hedonic regression shows a much higher price increase relative to the weighted hedonic regression.

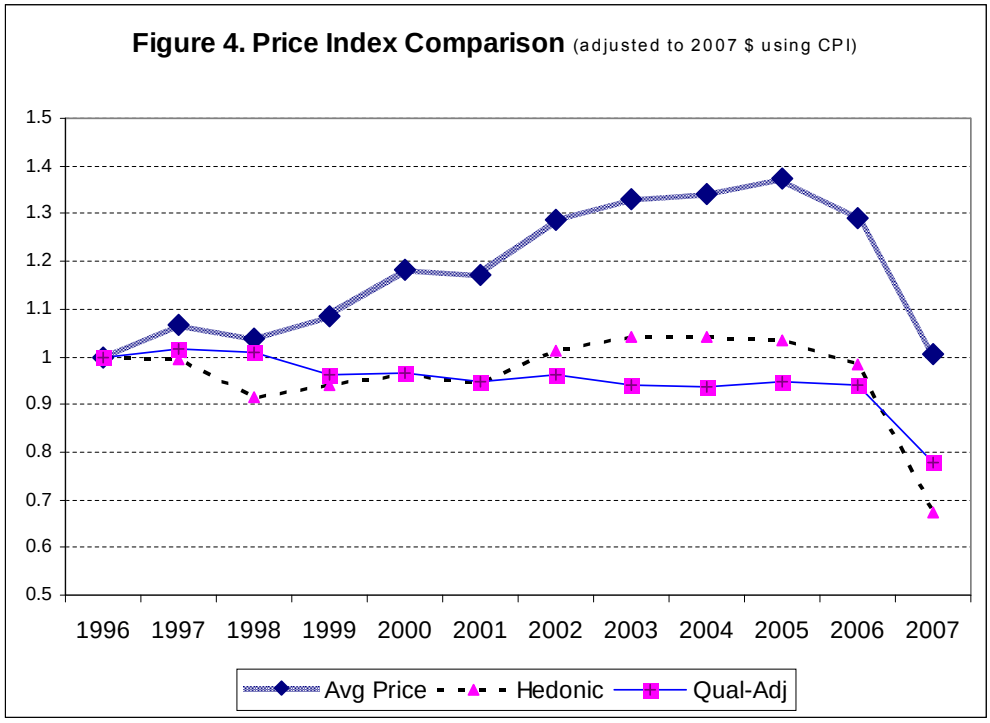
[44] To be more precise, the prices estimated from the Pakes hedonic approach are 1996: 1.000; 1997: .895; 1998: .853 1999: .964; 2000: 1.002; 2001: .937; 2002: 1.053; 2003: .964; 2004: .932; 2005: .933; 2006: .879; 2007: .679. One problem that may affect the stability of these estimates is that in my sample I observe a maximum of 30 products in a year, while Pakes (2003) examines around 200 products.

Table 7. Hedonic Price Regression on Log(Price)*

	Model 1		Model 2	
	Coef.	t	Coef.	t
LDL Cholesterol Reduction (% Amount)	0.032	(9.12)	0.117	(4.12)
LDL Cholesterol Reduction (% Amount)^2			0.001	(2.73)
HDL Cholesterol Reduction	-0.004	(-0.36)	-0.008	(-0.63)
Trigliciride Reduction Levels	0.000	(-0.03)	-0.004	(-1.07)
STATIN	0.252	(1.05)	-0.079	(-0.35)
YEAR 1997	0.008	(1.07)	-0.006	(-0.84)
YEAR 1998	-0.060	(-4.24)	-0.089	(-6.94)
YEAR 1999	-0.030	(-1.46)	-0.061	(-3.27)
YEAR 2000	0.000	(0.01)	-0.040	(-2.56)
YEAR 2001	-0.015	(-0.72)	-0.056	(-3.19)
YEAR 2002	0.054	(2.85)	0.014	(0.97)
YEAR 2003	0.080	(4.02)	0.042	(3.04)
YEAR 2004	0.072	(3.17)	0.041	(2.77)
YEAR 2005	0.053	(2.09)	0.035	(1.91)
YEAR 2006	-0.011	(-0.4)	-0.016	(-0.76)
YEAR 2007	-0.388	(-14.37)	-0.394	(-18.91)
Constant	-0.504	(-2.9)	-1.668	(-3.29)
	Frequency weighted		Frequency weighted	
Number of Observations	266		266	
Adj. R-Squared	0.419		0.455	

*Standard error estimates are clustered by year

Given the demand estimates and the hedonic estimates, all three price indices may be constructed. Figure 4 shows each of the three price indices: the average price, the hedonic price, and the quality-adjusted price derived from the demand estimates. While the average price increases by almost 37 percent from 1996 to 2005, the price index based on the demand estimates fell by 5 percent. The hedonic index is much closer to the quality-adjusted price index and increases by only 4 percent over this period, confirming the important role of quality in the determination of price in this market. There are clear differences across these indices pre-2005, but post 2005 all three indices show a large decrease in price after the introduction of the generic version of Zocor and Pravachol. Although the hedonic price and the price indices constructed from demand are relatively close in value, the hedonic approach may be problematic for other drug classes. First, for many other drug classes it may be difficult to find a dimension of quality that is as important as the reduction in LDL is for anti-cholesterol drugs. Second, as noted in Pakes (2003), the hedonic price may be significantly affected by the cost side of the market and is a reduced form approximation to individual welfare changes over time.



Several assumptions were made in constructing the quality-adjusted price index shown in Figure 4. To explore the importance of these assumptions, Table 8 presents alternative quality-adjusted price indices along with the average price, the hedonic price and a Laspeyres index. The Laspeyres index uses prior period expenditures to weigh price changes, similar to how price indices are currently constructed at the BLS.[45] The following are the different assumptions made for the four different estimates reported: (1) ignores moral hazard issues caused by private drug insurance and allows the trend variable and error, ξ_{jt}, to vary over time; (2) controls for moral hazard issues by removing effects of drug insurance, but allows the trend variable and error to vary over time (the result reported in Figure 4 above); (3) removes drug insurance effects and fixes the trend variable to its initial value, but allows the error to vary over time; (4) removes drug insurance effects and fixes the trend variable to its initial value and error terms are held constant over time.[46] The results show some variation in the price index (in some cases differing by as much as 13 percent), the differences appearing relatively minor when compared to the effect of not correctly measuring the value of new goods. That is, all four quality-adjusted price indices are much lower than the average price. Moreover, compared to the BLS price index, the quality adjusted prices are all 10-15 percentage points below the Laspeyres index in the last few years of the data.

[45] For the Laspeyres index, Generics and Branded versions of the same molecule are treated as an identical product, consistent with the current practice. In addition, the price index is computed using a geometric mean.

[46] The mean error terms that vary over time are set equal to zero.

The prices indices must be adjused slightly in 2007 to address the change in the questioning of respondents in 2007 that caused more individuals to report have a condition. In particular, this change would cause ξ_{jt} to fall for all drugs, since more individuals report a condition that previously did not. To address this issue I add 0.46 to ξ_{jt} for 2007 drugs, except for index (4). The value 0.46 was calculated as the average drop in ξ_{jt} from 2006 to 2007.

Each of the four indices differ substantially from the average price, but the large price reduction observed in 1997 using index (4) is quite different from indices (1), (2), and (3) that each show a small price increase followed by a gradual price decline. The reason for this difference is that index (4) removes the effects of ξ_{jt}, fixing the value of the drug, which implies that a drug like Lipitor that acquires greater share in later years, may have a larger initial effect on the price index. Although this initial difference is interesting, over time estimates shown in column (4) move closer to indices (1) through (3) and remain much lower than the average price over the entire period.[47]

Table 8. Price Index Comparison and Alternative Assumptions (adjusted to 2007 $ using CPI)

				Quality-Adjusted Price Indices			
Year	Avg Price	Laspeyres	Hedonic	(1)	(2)	(3)	(4)
1996	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1997	1.07	0.99	0.99	1.02	1.02	1.02	0.89
1998	1.04	0.98	0.92	1.01	1.01	1.02	0.87
1999	1.08	1.01	0.94	0.96	0.96	0.98	0.89
2000	1.18	1.01	0.96	0.96	0.97	0.99	0.89
2001	1.17	1.00	0.95	0.95	0.95	0.97	0.88
2002	1.29	1.04	1.01	0.96	0.96	0.99	0.92
2003	1.33	1.07	1.04	0.94	0.94	0.97	0.91
2004	1.34	1.07	1.04	0.93	0.94	0.97	0.90
2005	1.37	1.09	1.04	0.94	0.95	0.98	0.93
2006	1.29	1.08	0.98	0.93	0.94	0.98	0.93
2007	1.00	0.88	0.67	0.78	0.78	0.80	0.75
				With Insurance	Removing Insurance	Removing Insurance	Removing Insurance
				With Trend	With Trend	No Trend	No Trend
				With Error	With Error	With Error	No Error

In contrast to the results shown here, Nevo (2003) finds that quality-adjusted price indices vary greatly for breakfast cereals depending on whether, $Trend_t$, and, ξ_{jt}, are allowed to vary. A critical difference between Nevo's analysis and the market studied here is that unlike breakfast cereals, where innovations are relatively small, the innovations in prescription drug markets may be substantial. The results presented here show that the impact of addressing changes in product quality may be much more important in innovative industries, such as prescription drug markets, than the particular assumptions regarding unobserved demand characteristics. However, for typical consumer goods products where innovation is more limited it is likely that the assumptions regarding unobserved demand characteristics will be relatively more influential.

Disease-Specific Index. The above indices show a reduction in the quality-adjusted price in the aggregate, but the quality of the drug may also depend on the disease being treated. For instance, the demand estimates show that individuals with heart disease have a higher demand for anti-cholesterol drugs in general and also tend to prefer Zocor. This implies that those with heart disease may have different valuations for changes in the type of drug offerings in the market compared to individuals without heart disease, and the price for each group of individuals

[47] Note that even though ξ_{jt} and trend are fixed, I allow the age of the molecule to vary over time. As an alternative, not shown here, I fix the age of the Molecule to 10 years for all drugs. I obtain similar results, although the price decline reported in the index is slightly larger.

may be distinct. Therefore, it may be useful to construct a distinct price index for different populations.

As an illustration, Table 9 below helps illustrate how quality-adjusted prices may differ across populations with different diseases, using the hypethetical scenario of statin drugs entering the market in 1996 or in 2007.[48] While this exercise is purely based on a hypothetical that has not occurred (i.e. statins were not introduced in the market in 1996 or 2007), it is representative of the magnitude of the total impact of statin drug development on each type of health condition. Before presenting the quality adjusted price changes note that the presence of statin drugs make up a significant fraction of consumer welfare, accounting for 71.8 percent of total welfare in 1996 and 84.6 percent of total welfare in 2007.[49] These welfare impacts may be translated into a quality-adjusted price reduction of 21 percent in 1996 and a price reduction of 28 percent by 2007. (Given the large fraction of welfare attributed to the statin class, one might have expected an even larger price reduction. However, recall that this price index is a conservative estimate using current period prices and product characteristics, rather than base period products and characteristics.[50]) To determine the implied price reduction by disease type, the reduction in price is calculated for groups of individuals with heart disease and those without. The results show that the development of the statin class of drugs was more important for individuals with heart disease in both years, with a quality-adjusted price reduction that is 2.6 percentage points lower for individuals with heart disease in 1996 and 4 percentage points lower in 2007.

Table 9. Contribution of the Statin Class to Welfare

	1996	2007
Welfare from the Availability of Anti-cholesterol Drugs (in billions 2007 $)	$1.07	$6.81
Welfare from the Availability of Statin Drugs (in billions 2007 $)	$0.77	$5.76
Fraction of Welfare From Statins	**71.8%**	**84.6%**
Hypothetical Reduction in Quality-Adjusted Price From Introduction of Statins:		
Overall Decrease	-21.0%	-27.7%
Decrease for Those with Heart Disease	-22.8%	-30.5%
Decrease for Those without Heart Disease	-20.2%	-26.5%

Table 9 shows that a disease-specific index may be calculated for anti-cholesterol drugs and that one may find measurable differences in price changes depending on an individual's health condition. For researchers attempting to track the cost of disease treatment, the above indices may be used to discount expenditures on anti-cholesterol treatments for two distinct populations, those with heart disease and those without. A similar methodology may be applied across alternative drug classes and disaggregation of price indices into alternative disease or age categories

[48] These estimates differ from the welfare estimates reported in Table 5 because they are based on consumer welfare estimates with insurance effects removed.

[49] These figures are calculated by estimating welfare when the Statin drugs are available compared to welfare estimates when the Statin drugs are not avaialable.

[50] Given the large fraction of welfare attributed to the statin class, one might have expected an even larger price reduction. However, recall that this price index is a conservative estimates using current period prices and product characteristics, rather than base period products and characteristics. The price change may be considerably larger if one were interested in the price change using the base period products consisting of non-statins.

is also feasible.

8 Conclusion

To assess the importance of innovation or the productivity in the health care sector, it is essential to have an understanding of how new technologies affect the quality of treatment and individual welfare. This paper focuses on measuring the impact of innovations in the market for high cholesterol treatments that has experienced both several new product introductions and a large increase in overall expenditures.

The impact of innovation on welfare is measured using a price index that holds the quality of anti-cholesterol drug treatments fixed over time. To estimate a quality-adjusted price, I first estimate the market demand for anti-cholesterol drugs using nationally-representative micro level data. The demand model is then used to estimate a price index that holds the quality of treatment constant over time. I find that this price index fell by 5 percent from 1996 to 2005. This contrasts sharply with the average price that has increased by 37 percent over this period. The importance of controlling for quality is confirmed using a hedonic index that shows a price increase of just 4 percent, which is far below the average price and much closer in value to the quality-adjusted index.

Prior research by Nevo (2003) finds that price indices constructed from estimated demand systems vary greatly for breakfast cereals depending on the particular assumptions made regarding unobserved demand characteristics (i.e. whether unobserved demand represents changing taste or an unobserved change in the product attributes). I also find that these assumptions affect quality-adjusted prices in the market for anti-cholesterol drugs. However, I find that assumptions regarding demand unobservables and trends have a relatively small effect on the price index derived from the demand model in comparison to the very large price increases when using an average price. A likely reason for this difference is that accounting for product quality is much more important in innovative industries, such as prescription drug markets, while the unobserved demand characteristics will be relatively more influential in determining quality-adjusted price changes for typical consumer good products where innovation is limited. Therefore, while policy-makers should remain cautious in applying market demand estimates to construct price indices for a broad range of products, it may be important to estimate price indices from estimated demand in innovative markets, such as prescription drugs, where accounting for quality changes is likely to be critical for obtaining a meaningful price index.

An important advantage of deriving a price index from micro demand estimates is that one can allow the value of medication to depend on the health condition of the individual being treated. Although most individuals with high cholesterol benefit from new cholesterol treatments over the period of study, I find that those with high cholesterol and heart disease, benefit more from the introduction of statin drugs compared to those without heart disease, and this has a measurable impact on quality-adjusted prices. Evaluating the impact of new pharmaceuticals on a disease specific price may be important for the broader agenda of evaluating the benefits of newly introduced technologies on the cost of disease treatment.

There are at least three areas where the research presented here may be extended. First, the approach used in this paper may be applied to other pharmaceutical categories to determine how new technologies impact the cost of disease treatment for other diseases. Second, alternative models that include random coefficients at the micro level are more computationally intensive, but may provide additional insight into the value of new products. It may be useful to see if alternative approaches to estimating demand produce measurably different quality-adjusted price indices. Third, in the literature review of this paper I highlight how the market value of a drug may differ from the value implied by looking at health expenditures and health outcomes. Our understanding of innovation could be improved through future research analyzing whether these methodologies produce different results, and if so, why.

9 Bibliography

References

[1] Aizcorbe, Ana, and Nicole Nestoriak, (2010), "Price Indexes for Drugs: A Review of the Issues", Manuscript

[2] Aizcorbe, Ana, and Nicole Nestoriak, (2010), "Changing Mix of Medical Care Services: Stylized Facts and Implications for Price Indexes", Manuscript

[3] Berry, Steven and James Levinsohn and Ariel Pakes, (1995), "Automobile Prices in Market Equilibrium", Econometrica, 63(4) pgs 841-890

[4] Berry, Steven, James Levinsohn, and Ariel Pakes, (2004), "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market", Journal of Political Economy 112(1) pgs 68-105

[5] Berndt, Ernst, Linda Bui, David Reiley, and Glen Urban, (1995), "Information, Marketing, and Pricing in the U.S. Antiulcer Drug Market", American Economic Review, 85(2) pgs 100-105

[6] Berndt, Ernst, David Cutler, Richard Frank, Zvi Griliches, Joseph Newhouse, and Jack Triplet, (2000), "Medical Care Prices and Output", Chapter 3, Handbook of Health Economics, Volume 1

[7] Berndt, Ernst, Anupa Bir, Susan H. Busch, Richard Frank, Sharon-Lise T. Normand, (2002), "The Medical Treatment of Depression, 1991-1996: productive inefficiency, expected outcome variations, and price indices", Journal of Health Economics

[8] Bresnahan, Timothy and Robert Gordon, (1997), "The Economics of New Goods", NBER Book Series Studies in Income and Wealth

[9] Chintagunta, Pradeep, Jean-Pierre Dube, and Kim Yong Goh, (2005), "Beyond the endogeneity bias : The effect of unmeasured brand characteristics on household-level brand choice models", Management Science, 51(5), pgs 832-849

[10] Cleanthous, Paris, (2004), "Patient Welfare Implications of Innovation in the U.S. Antidepressant Market", Working Paper

[11] Cutler, David, Mark McClellan, Joseph Newhouse, and Dalia Remler, (1998), "Are Medical Prices Declining? Evidence From Heart Attack Treatments", Quarterly Journal of Economics, 63(4) pgs 691-1024

[12] Culter, David and Mark McClellan, (2001), "Is Technological Change in Medicine Worth It?", Health Affairs, 20(5) pgs 11-29

[13] Congressional Budget Office, (2008), "Technological Change and the Growth of Health Care Spending", The Congress of the United States, Congressional Budget Office. January

[14] Congressional Budget Office, (2006), "Research and Development in the Pharmaceutical Industry", The Congress of the United States, Congressional Budget Office. October

[15] Epstein, L and J. Zhang, (2001), "Subjective Probabilities on Subjectively Unambiguous Events", Econometrica, 69 pgs 265-306

[16] Ellison, Fisher, Ian Cockburn, Zvi Griliches, and Jerry Hausman, (1996), "Characteristics of Demand for Pharmaceutical Products: an Examination of four Cephalosporius", Rand Journal of Economics, 28(3) pgs 426-446

[17] Gaynor, Martin and William Vogt, (2003), "Competition Among Hospitals", RAND Journal of Economics, 34(4) pgs 764-785

[18] Goolsbee, Austan, and Amil Petrin, (2004), "A Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," Econometrica, 72(2) pgs 351-381

[19] Gerberding, Julie L., Michael Leavitt, and Edward Sondick, (2009), "Health, United States, 2009" DHHS Publication No. 2005-1232

[20] Heart Protection Study, (2010), "Statin Cost-Effectiveness in the United States for People at Different Vascular Risk Levels", Circulation: Cardiovascular Quality and Outcomes, 2 pgs 65-72

[21] Kennan, John, (1989), "Simultaneous Equation Bias in Disaggregated Econometric Models", Review of Economic Studies, 56(1) pgs 151-156

[22] Lucarelli, Claudio and Sean Nicholson, (2009), "A Quality-Adjusted Price Index for Colorectal Cancer Drugs", NBER Working Paper Series, Working Paper 15174

[23] Mcfadden, D., (1981) "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden, Structural Analysis of Discrete Data, Cambridge, MA: The MIT Press

[24] National Cholesterol Education Program, (2001) Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). National Institutes of Health, National Heart, Lung, and Blood Institute

[25] Nevo, Aviv, (2003), "New Products, Quality Changes, and Welfare Measures Computed from Estimated Demand Systems" The Review of Economics and Statistics, 85(2) pgs 266-275

[26] Pakes, Ariel, (2003), "A Reconsideration of Hedonic Price Indexes with an Application to PC's", American Economic Review, 93(5) pgs 1578-1596

[27] Pakes, Ariel, Steven Berry, and James Levinsohn, (1993), "Applications and Limitations of Some Recent Advances in Empirical I.O.: Price Indexes and Analysis of Environmental Change." American Economic Review, 83(2) pgs 240-46

[28] Petrin, Amil, (2002), "Quantifying the Benefits of New Products: The Case of the Minivan", Journal of Political Economy, 110(4) pgs 705-729

[29] Romeo, Charles, (2010), "Filling Out the Instrument Set in Mixed Logit Demand Systems for Aggregate Data", U.S. Department of Justice Working Paper

[30] Scandinavian SimvaStatin Survival Group, (1994), "Randomized Trial of Cholesterol Lowering in 4,444 Patients with Coronary Heart Disase: the Scandanavian SimvaStatin Survival Study (4S), Lancet, 344 pgs 1383-9

[31] Smith, Sheila, Stephen Heffler and Mark Freeland, (2000), "The Impact of Technological Change on Health Care Cost Increases", Working Paper

[32] Smith, Sheila, Joseph Newhouse, and Mark Freeland, (2009), "Income, Insurance, and Technology: Why Does Health Spending Outpace Economic Growth", Health Affairs, 28(5) pgs 1276-84

[33] U.S. Department of Health And Human Services, (2000), Report to the President: Prescription Drug Coverage, Spending, and Prices

[34] Trajtenberg, Manuel, (1990), "Product Innovation, Price Indices, and the (Mis)Measurement of Economic Performance", National Bureau of Economic Research, Working Paper No. 3261

[35] Villas-Boas, J. and Russell Winer, (1999), "Endogneity in Brand Choice Models", Management Science, 45 pgs 1324-1338

[36] Ward, S., M. Lloyd Jones, A Pandor, M Holmes, R Ara, A Ryan, W Yeo, and N Payne, (2007), "A Systematic Review and Economic Evaluation of Statins for the Prevention of Coronary Events", Health Technology Assessment, 11(14)

[37] Wosinska, Marta, (2005), "Direct-to-consumer Advertising and Drug Therapy Compliance", Journal of Marketing Research, 42(3), pgs 323-332

[38] World Health Organization, World Health Report 2002: Reducing risks, promoting health life, (2002), Geneva

10 Appendix

10.1 Drug Characteristics

Table A1. Drug Characteristics: Patent Holder, Class, Strength, Approval Date, Effectiveness

BRAND NAME	ACTIVE INGREDIENT	PATENT HOLDER	DRUG CLASS	STRENGTH	APPROVAL DATE	In Sample	TOT CHOL	LDL	HDL	TRI
LIPITOR	ATORVASTATIN	PFIZER	STATIN	10	12/17/1996	Yes	-29	-39	6	-19
LIPITOR	ATORVASTATIN	PFIZER	STATIN	20	12/17/1996	Yes	-33	-43	9	-26
LIPITOR	ATORVASTATIN	PFIZER	STATIN	40	12/17/1996	Yes	-37	-50	6	-29
LIPITOR	ATORVASTATIN	PFIZER	STATIN	80	12/17/1996	No	-45	-60	5	-37
BAYCOL	CERIVASTATIN	BAYER PHARMS	STATIN	0.3	6/26/1997	Yes	-22	-31	8	-16
BAYCOL	CERIVASTATIN	BAYER PHARMS	STATIN	0.4	6/26/1997	Yes	-24	-34	7	-16
BAYCOL	CERIVASTATIN	BAYER PHARMS	STATIN	0.8	6/26/1997	No	-30	-42	9	-22
LESCOL	FLUVASTATIN	NOVARTIS	STATIN	20	12/31/1993	Yes	-17	-22	3	-12
LESCOL	FLUVASTATIN	NOVARTIS	STATIN	40	12/31/1993	Yes	-19	-25	4	-14
LESCOL	FLUVASTATIN	NOVARTIS	STATIN	80	12/31/1993	No	-25	-36	6	-18
LOVASTATIN	LOVASTATIN	GENERIC	STATIN	10	12/17/2001	No	-16	-21	5	-10
LOVASTATIN	LOVASTATIN	GENERIC	STATIN	20	12/17/2001	Yes	-17	-24	7	-10
LOVASTATIN	LOVASTATIN	GENERIC	STATIN	40	12/17/2001	Yes	-22	-30	7	-14
MEVACOR	LOVASTATIN	MERCK	STATIN	10	8/31/1987	No	-16	-21	5	-10
MEVACOR	LOVASTATIN	MERCK	STATIN	20	8/31/1987	Yes	-17	-24	7	-10
MEVACOR	LOVASTATIN	MERCK	STATIN	40	8/31/1987	Yes	-22	-30	7	-14
MEVACOR	LOVASTATIN	MERCK	STATIN	60	8/31/1987	No	-29	-40	10	-19
PRAVOCOL	PRAVASTATIN	BRISTOL-MEYER SQUIB	STATIN	10	10/31/1991	No	-16	-22	7	-15
PRAVOCOL	PRAVASTATIN	BRISTOL-MEYER SQUIB	STATIN	20	10/31/1991	Yes	-24	-32	2	-11
PRAVOCOL	PRAVASTATIN	BRISTOL-MEYER SQUIB	STATIN	40	10/31/1991	Yes	-25	-34	12	-24
PRAVOCOL	PRAVASTATIN	BRISTOL-MEYER SQUIB	STATIN	80	10/31/1991	No	-27	-37	3	-19
PRAVASTATIN	PRAVASTATIN	GENERIC	STATIN	10	4/24/2006	No	-16	-22	7	-15
PRAVASTATIN	PRAVASTATIN	GENERIC	STATIN	20	4/24/2006	Yes	-24	-32	2	-11
PRAVASTATIN	PRAVASTATIN	GENERIC	STATIN	40	4/24/2006	Yes	-25	-34	12	-24
PRAVASTATIN	PRAVASTATIN	GENERIC	STATIN	80	4/24/2006	No	-27	-37	3	-19
CRESTOR	ROSUVASTATIN	ASTRAZENECA	STATIN	20	8/12/2003	Yes	-40	-55	8	-23
CRESTOR	ROSUVASTATIN	ASTRAZENECA	STATIN	40	8/12/2003	No	-46	-63	10	-28
CRESTOR	ROSUVASTATIN	ASTRAZENECA	STATIN	5	8/12/2003	No	-33	-44	13	-35
CRESTOR	ROSUVASTATIN	ASTRAZENECA	STATIN	10	8/12/2003	Yes	-36	-52	14	-10
ZOCOR	SIMVASTATIN	MERCK	STATIN	10	12/23/1991	Yes	-23	-30	12	-15
ZOCOR	SIMVASTATIN	MERCK	STATIN	20	12/23/1991	Yes	-28	-38	8	-19
ZOCOR	SIMVASTATIN	MERCK	STATIN	40	12/23/1991	Yes	-31	-41	9	-18
ZOCOR	SIMVASTATIN	MERCK	STATIN	5	12/23/1991	No	-19	-25	10	-12
ZOCOR	SIMVASTATIN	MERCK	STATIN	80	12/23/1991	No	-36	-47	8	-24
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	10	12/20/2006	Yes	-23	-30	12	-15
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	20	12/20/2006	Yes	-28	-38	8	-19
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	40	12/20/2006	Yes	-31	-41	9	-18
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	5	12/20/2006	No	-19	-25	10	-12
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	80	12/20/2006	No	-36	-47	8	-24
VYTORIN	EZTIMBE/SIMVASTATIN	MSP SINGAPORE CO LLC	COMBO	10	7/23/2004	No	-31	-45	8	-23
VYTORIN	EZTIMBE/SIMVASTATIN	MSP SINGAPORE CO LLC	COMBO	20	7/23/2004	Yes	-36	-52	10	-24
VYTORIN	EZTIMBE/SIMVASTATIN	MSP SINGAPORE CO LLC	COMBO	40	7/23/2004	Yes	-39	-55	6	-23
VYTORIN	EZTIMBE/SIMVASTATIN	MSP SINGAPORE CO LLC	COMBO	80	7/23/2004	No	-43	-60	6	-31
ADVICOR	LOVASTATIN/NIACIN	ABBOTT LABORATORIES	COMBO	20	12/17/2001	Yes	-21	-30	20	-32
CLOFIBRATE	CLOFIBRATE	GENERIC	FIBRIC ACID	500	2/8/1967	Yes	-18	-21	11	-29
FENOFIBRATE	FENOFIBRATE	FENOFIBRATE	FIBRIC ACID	145	12/31/1993	Yes	-18	-21	11	-29
FENOFIBRATE	FENOFIBRATE	FENOFIBRATE	FIBRIC ACID	145	5/13/2005	Yes	-18	-21	11	-29
GENERIC	GEMFIBROZIL	GENERIC	FIBRIC ACID	600	9/29/1995	Yes	-18	-21	11	-29
LOPID	GEMFIBROZIL	PFIZER	FIBRIC ACID	600	12/21/1981	Yes	-18	-21	11	-29
NIACIN	NIACIN	GENERIC	NICOTINIC ACID	500	1/1/1970	Yes	-5	-9	15	-11
NIASPAN	NIACIN	KOS	NICOTINIC ACID	500	7/28/1997	Yes	-5	-9	15	-11
ZETIA	EZTIMBE	MSP SINGAPORE CO LLC	OTHER	10	10/25/2002	Yes	-14	-20	4	-5
CHOLESTYRAMINE	CHOLESTYRAMINE	BRISTOL-MEYER SQUIB	SEQUESTRANTS	5.7	8/15/1996	Yes	-17	-23	4	0
GENERIC	CHOLESTYRAMINE	GENERIC	SEQUESTRANTS	5.7	8/3/1973	Yes	-17	-23	4	0
WELCHOL	COLESEVELAM	SANKOYO	SEQUESTRANTS	625	5/26/2000	Yes	-10	-18	3	-9
GENERIC	COLESTIPOL	GENERIC	SEQUESTRANTS	1	5/2/2006	Yes	-17	-23	4	0
COLESTID	COLESTIPOL	PHARMACIA AND UPJOHN	SEQUESTRANTS	1	4/4/1977	Yes	-17	-23	4	0

10.2 Estimates

Table A2. Probit of Risk Factors on the Decision to Take A Drug

	dF/dx	z
Has High Cholesterol	0.506	(65.85)
Atherosclerotic Condition	0.073	(10.02)
Has Diabetes	0.057	(8.63)
Has Hypertension	0.047	(8)
Age>=40	0.042	(2.24)
Age	0.020	(3.66)
Age^2	0.000	(0.89)
Age^3	0.000	(-4.1)
Male	0.130	(4.87)
Age*Male	-0.001	(-2.75)
Perceived Health is Good	-0.032	(-4.26)
Number of Observations	106,510	
Pseudo R^2	0.187	

Table A3. (Table 4. Continued). First Stage Results from Conditional Logit Estimation

Variable	Coef.	z-stat
Heart Disease*Lipitor	-0.173	(-3.17)
Heart Disease*Baycol	-0.844	(-3.42)
Heart Disease*Lescol	-0.448	(-3.66)
Heart Disease*Mevacor (or Generic Mev)	-0.428	(-4.28)
Heart Disease*Pravachol (or Generic Prav)	-0.099	(-1.21)
Heart Disease*Crestor	-0.093	(-0.75)
Heart Disease*Advicor	-0.162	(-0.42)
Heart Disease*Fibric Acid Derivative	-0.300	(-2.98)
Heart Disease*Nictonic Acid	0.035	(0.19)
Heart Disease*Sequestrant	-0.765	(-3.21)
Heart Disease*Zetia	-0.271	(-1.71)
Heart Disease*Generic	-0.064	(-0.79)
High Chol*Lipitor	0.210	(2.5)
High Chol*Baycol	-0.470	(-1.46)
High Chol*Lescol	-0.036	(-0.22)
High Chol*Mevacor (or Generic Mev)	0.022	(0.16)
High Chol*Pravachol (or Generic Prav)	0.030	(0.26)
High Chol*Crestor	0.282	(1.45)
High Chol*Advicor	-0.116	(-0.23)
High Chol*Fibric Acid Derivative	0.173	(1.07)
High Chol*Nictonic Acid	-0.013	(-0.06)
High Chol*Sequestrant	0.189	(0.53)
High Chol*Zetia	0.092	(0.36)
High Chol*Generic	0.317	(2.17)
Age*Lipitor	-0.005	(-2.79)
Age*Baycol	0.011	(1.54)
Age*Lescol	0.011	(2.85)
Age*Mevacor (or Generic Mev)	-0.004	(-1.5)
Age*Pravachol (or Generic Prav)	-0.001	(-0.4)
Age*Crestor	-0.009	(-2.35)
Age*Advicor	-0.019	(-1.6)
Age*Fibric Acid Derivative	-0.020	(-5.77)
Age*Nictonic Acid	-0.003	(-0.61)
Age*Sequestrant	0.011	(1.45)
Age*Zetia	0.011	(2.18)
Age*Generic	0.008	(2.94)

Table A4. First Stage Estimation for IV Model 1

Variable	Coef.	
markup	-11.32	(-0.19)
markup*generic	2251.99	(4.07)
demand	0.00	(2.11)
demand*generic	0.00	(-5.94)
Number of Observations	266	
R-Squared	0.925	

10.3 Demographics as Instruments in a Linear Demand Model using Micro Data

Using econometric theory Kennan (1989) shows that in some specialized settings observing individual data may be helpful in identifying demand; but more generally, in many commonly observed market scenarios, he shows that researchers cannot assume that micro data solves the endogeneity problem. Many empirical papers have confirmed his result by providing several examples where micro data does not correct the endogeneity problem (e.g. Villas-Boas and Winer (1999), Gaynor and Vogt (2003), Goolsbee and Petrin (2004), and Chintagunta et al (2005)). In cases where micro data does not solve the endogeneity problem, Kennan (footnote 5) hints at another potential benefit of using micro data: that the aggregate demographics in the market may be used as an instruments for price when individual demographics are included in the model.[51] However, he does not formally show when or why aggregate demographics are valid instruments. This section presents a brief proof that shows what assumptions are necessary for aggregate demographics to be valid instruments and provides a short discussion of why these instruments are different from those commonly used in the literature.

Rather than focus on a discrete choice model, I analyze a simple linear demand model similar to that analyzed in Kennan where it is easier to prove the exogeneity of aggregate demographics. Consider the following demand model:

$$Y_{it} = \alpha p_t + \beta z_{it} + \omega_{it} \tag{5}$$

where Y_{it} is the quantity demanded for individual i in market t where there are a total of T markets and N consumers in each market. Let p_t be the price of the product in market t that is set simultaneously by a single producer in the market and is an endogenous variable; let z_{it} be the demographics for individual i in market t (e.g. age, sex, education, income, or size of household). The parameters to be estimated include α and β and the error term is ω_{it}. Finally let the aggregate demographic variable, Z_t, simply be the mean demographics across the N individuals in

[51] Gaynor and Vogt exploit this basic idea, but they provide little additional discussion regarding when or how demographics may be used as an instrument.

market t:

$$Z_t = \frac{\sum_{i=1}^{N_t} z_{it}}{N_t}$$

Alternatively, to construct an instrument analagous to the discrete choice estimation, in a linear model one may use the aggregate average utility as an instrument, $Z_t = \frac{\sum_{i=1}^{N_t} \beta z_{it}}{N_t}$.[52]

There are three necessary assumptions for the variable Z_t to be a valid instrument. First, the individual's own demographics need to be exogenous so that $E(z_{it}\omega_{it}) = 0$. Note that this assumption is also necessary for the demographics to enter the above model and is a commonly made assumption in most demand models. Second, the individual's own demand for a product cannot be dependant on the demographics of other individuals in the market. So for individual $j \neq i$ it is assumed that $E(z_{it}\omega_{jt}) = 0$. This second assumption should hold if there are no network effects that might cause the demographics of other individuals in the market to affect one's own utility. In addition to the above assumption, it is also necessary for the instrument to be correlated with the market price of the good, p_t, which should occur if the demographics have an impact on demand and there is variation in demographics across markets. Given these three assumptions the consistency of an estimator that uses Z_t as an instrument follows a simple proof, since it is only necessary to show that $plim_{NT\to\infty} \frac{\sum_{t=1}^{T} \omega_{it} z_t}{NT} \to 0$.

proof. To show that $plim_{NT\to\infty} \frac{\sum_{t=1}^{T} \omega_{it} Z_t}{NT} \to 0$, I begin by substituting the value of Z_t and expanding the summation.

$$\begin{aligned} plim_{NT\to\infty} \frac{\omega_{it} Z_t}{NT} &= plim_{NT\to\infty} \frac{\sum_{t=1}^{T} \omega_{it} \left(\frac{\sum_{i=1}^{N} Z_i}{N} \right)}{NT} == plim_{NT\to\infty} \frac{\sum_{t=1}^{T} \omega_{it} \left(\sum_{i=1}^{N} z_i \right)}{N^2 T} \\ &= plim_{NT\to\infty} \frac{\sum_{t=1}^{T} \omega_{it} z_i + \omega_{it} \left(\sum_{j\neq i}^{N} z_j \right)}{N^2 T} = plim_{NT\to\infty} \frac{\sum_{t=1}^{T} \omega_{it} z_i}{N^2 T} + plim_{NT\to\infty} \frac{\sum_{t=1}^{T} \omega_{it} \left(\sum_{j\neq i}^{N} z_j \right)}{N^2 T} \end{aligned}$$

Given that it is assumed that $E(z_{it}\omega_{it}) = 0$ and for individual $j \neq i$ it is also assumed $E(z_{it}\omega_{jt}) = 0$, then it follows that $plim \frac{\sum_{t=1}^{T} \omega_{it} Z_t}{NT} \to 0$.

This instrumenting strategy relies on an asymmetry that is present in many markets. That is, producers price to the entire market, so that only market wide demographics are relevant, not a particular individual. In contrast,

[52] The effect of demographics on individual utility, βz_{it}, may be estimated by using a first stage estimation of 5 that includes period t dummy variables, δ_t, instead of p_t.

an individual's consumption decision is affected only by her characteristics and is unaffected by the characteristics of other consumers in the market.[53]

This approach may be compared to the IV strategy of using rival product characteristics to instrument for price because they both related to consumer preferences rather than shifts in underlying costs. However, it is important to note that instruments using demographics are different in at least two important ways: First, demographics are a valid instrument if the unobserved demand shock have a proportional effect on the population and firms are unable to alter products in a way to target a particular demographic (i.e. These instruments work well if firms have a limited ability to target certain demographic populations). Second, a bias typically arises because there is an unobserved attribute that the econometrician does not observe that is correlated with price. The key advantage of having consumer level data is that the consumers do observe the unobserved demand changes, even if the econometrician does not. Therefore, the econometrician may allow the consumers to respond to unobserved attributes which may correct for a potential bias.

To see how constructing instruments from demographics differs from the attributes of other products, it is useful to consider an example where the demographics are not exogenous so that $E(z_{it}\omega_{jt}) \neq 0$. For instance, if a firm is able to change a non-price component of a product to appeal to a particular demographic in a particular market. In this case, there will be an unobserved component of the product characteristic that will be correlated with the individual's valuation of a product so that $E(z_{it}\omega_{jt}) \neq 0$. For example, consider a hypothetical market for breakfast cereals where there is only a single producer and there are two versions of the cereal, one version that appeals to older adults and one that appeals to younger adults. Also suppose that a manufacturer only distributes one version of the cereal in each market and that the econometrician does not observe which version is introduced. In this case, one should expect the cereal manufacturer to distribute the version that appeals to older adults in markets where the average age is higher. Since the econometrician does not observe the version of the product, the assumption $E(z_{it}\omega_{jt})$ will be violated. This problem is analogous to problems that arise when one assumes that rival product characteristics are exogenous, when they may, in fact, be chosen strategically.

One important fix to this problem can be implemented whether demographic information is available or not. Namely, the inclusion of product fixed effects accounts for the product characteristics associated with the creation and introduction of the product that attempts to appeal to a particular demographic. (For the linear demand model 5 this implies including a constant). The inclusion of product fixed effects should be sufficient in cases where it is difficult for the manufacturer to customize their product across markets. In addition to including a product fixed effect, if individual information is available, one can allow the individual's value of the product to be different across markets. That is, estimate a model that allows for demographic information to have a unique effect in each market, so that individuals can flexibly respond to the quality of the product that is available in their market. In the above model this implies allowing for a different β in each market:

[53] Of course, this instrumenting strategy fails if producers can perfectly price discriminate, since each individual would have a unique price.

$Y_{it} = \alpha p_t + \beta_1 z_{i1} + \beta_2 z_{i2} + ... + \beta_T z_{iT} + \omega_{it}$

To translate this correction to the example of the breakfast cereals, this implies allowing age to have a unique effect in each market. In this case, even though the econometrician does not observe which version of the breakfast cereal was introduced in the market, it is sufficient that the individual does. While this particular fix may address the problem that unobserved product characteristics may vary across markets and be correlated with the demographics, the data requirements are significantly greater and may also be computationally burdensome, depending on the number of markets. More practically, even if it is not feasible to allow demographics to a have a unique effect in each market, the econometrician should allow for some flexibility in how consumers respond to product characteristics across markets to reduce the potential for endogeneity bias. For example, the econometrician may allow for a different parameter, β, in markets where the average age is older. The econometrician may also allow for greater flexibility in how consumer demographics are interacted with the product characteristics.

To allow for greater flexibility in the anti-cholesterol drug demand model in this paper, I interact the consumer characteristics with molecule dummies to allow for a distinct reaction to each molecule depending on an individual's health condition. However, more generally, one might also consider interacting demographics with unobserved mean utility, ξ_{jt}, to allow consumers to have a distinct response to changes in unobserved product attributes (e.g. $\gamma age \cdot \xi_{jt}$), although this may increase the computational difficulty of the estimation procedure it may reduce the potential for endogeneity in cases where the demographics are used as instruments.